

FIGURE 1

Phylogenetic Classification of HIV-1

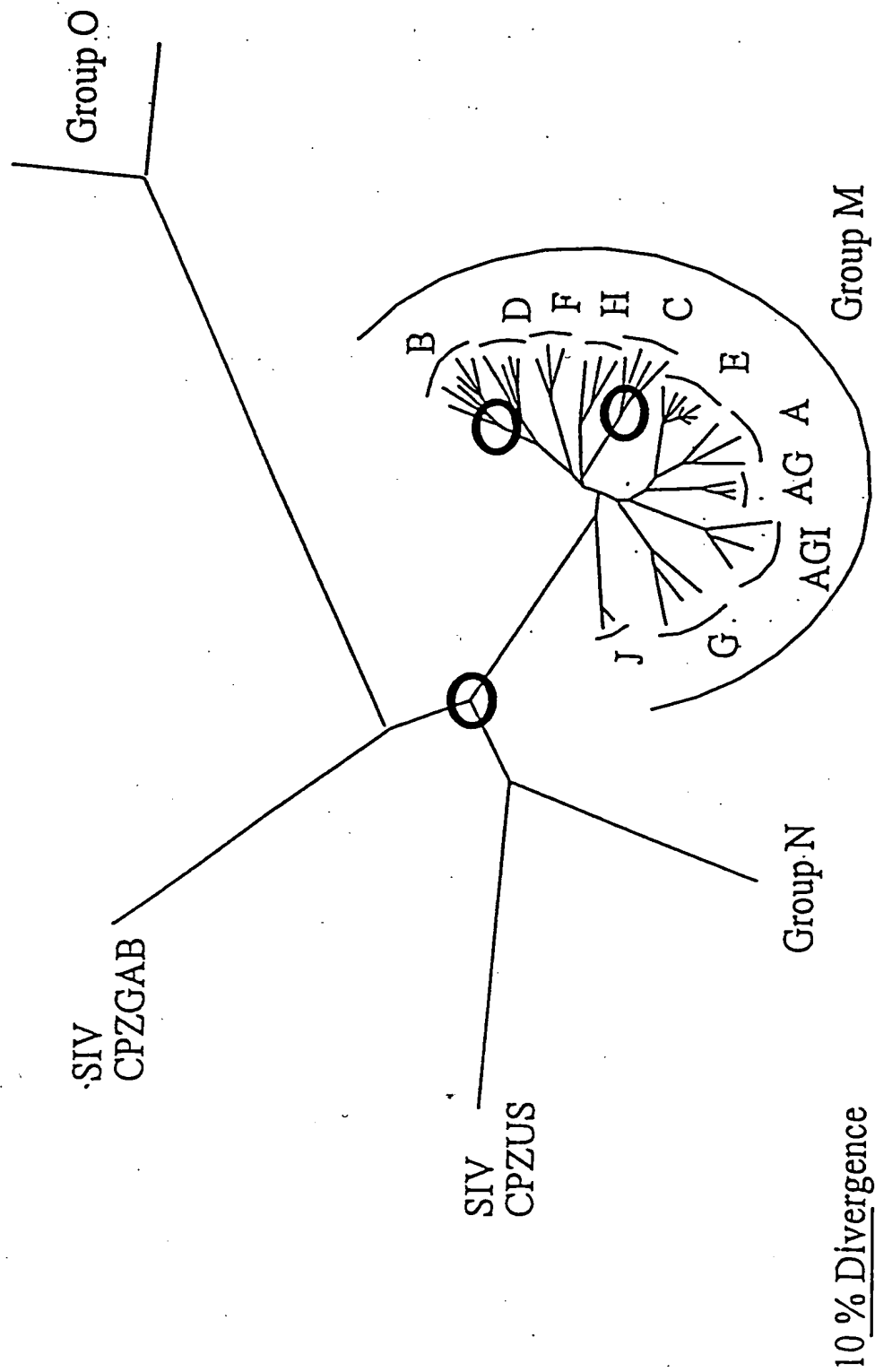


Figure 2

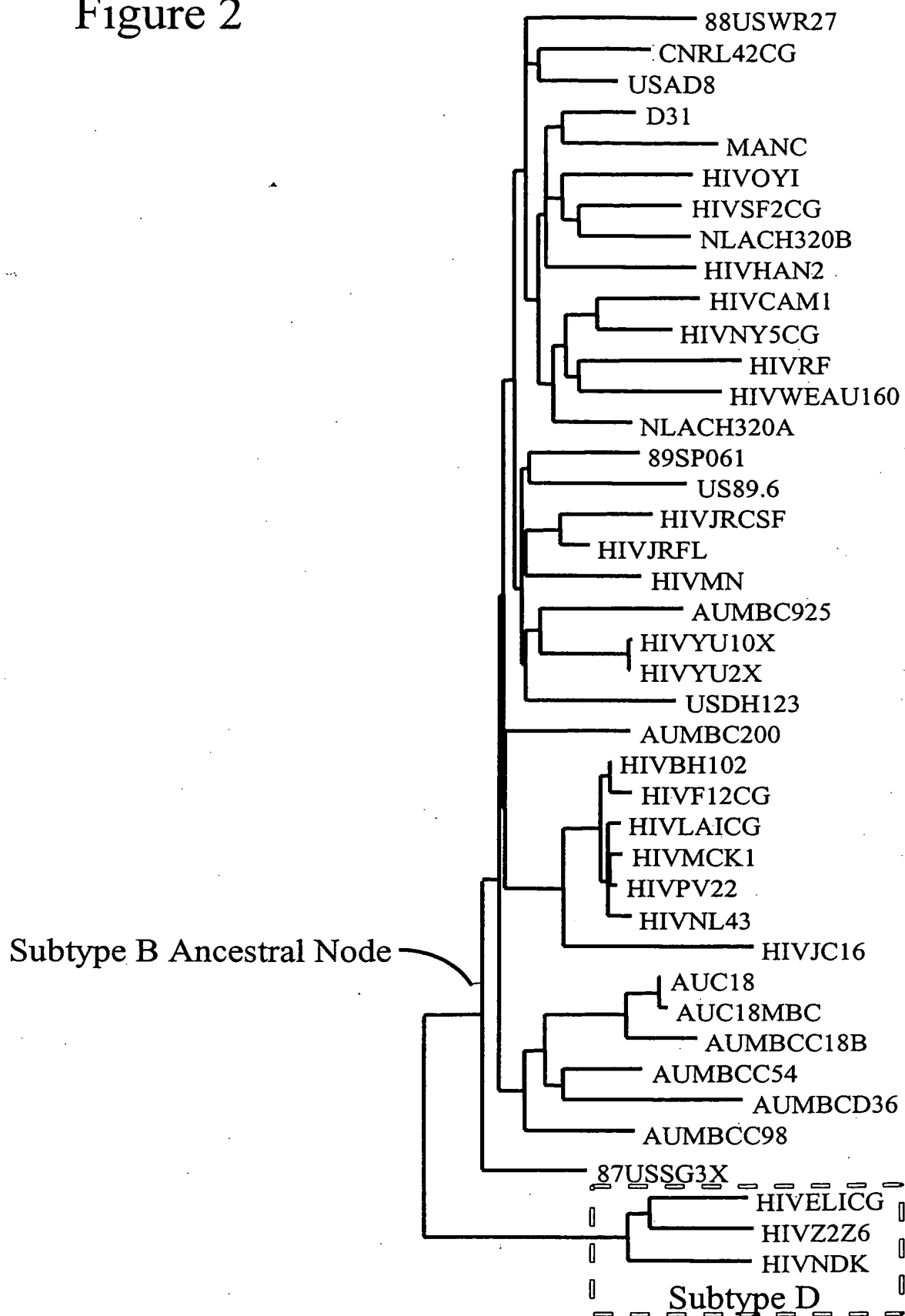


Figure 3

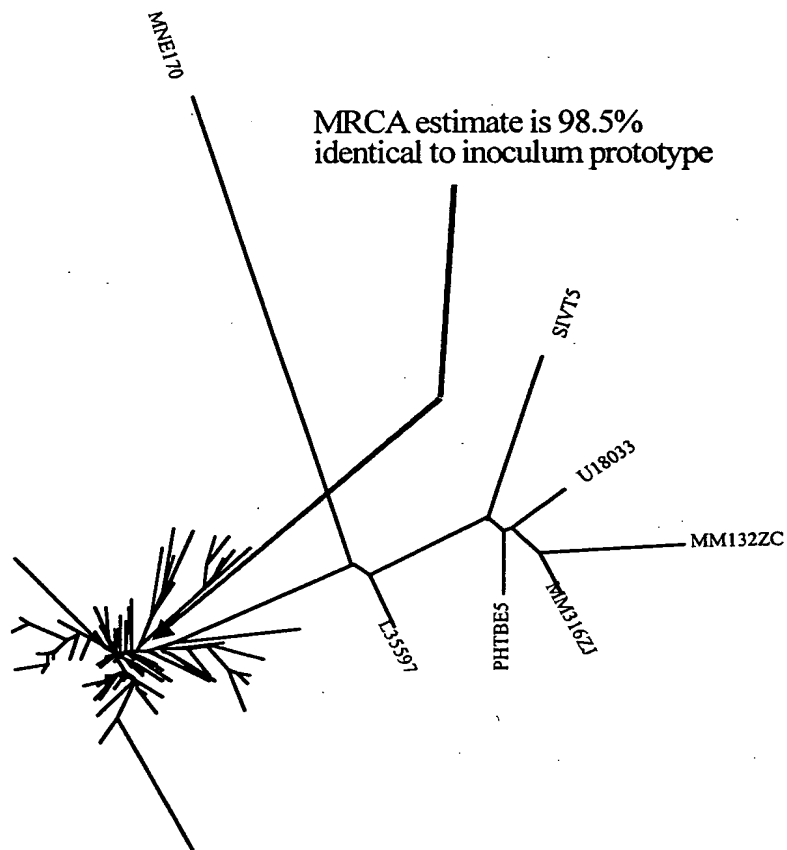


Figure 4

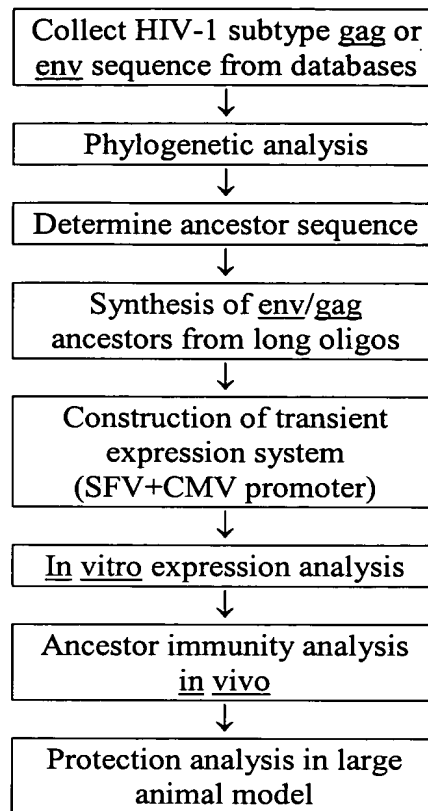


Figure 5

MP
Reconstruction



ML
Reconstruction

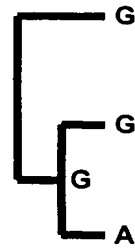


Figure 6

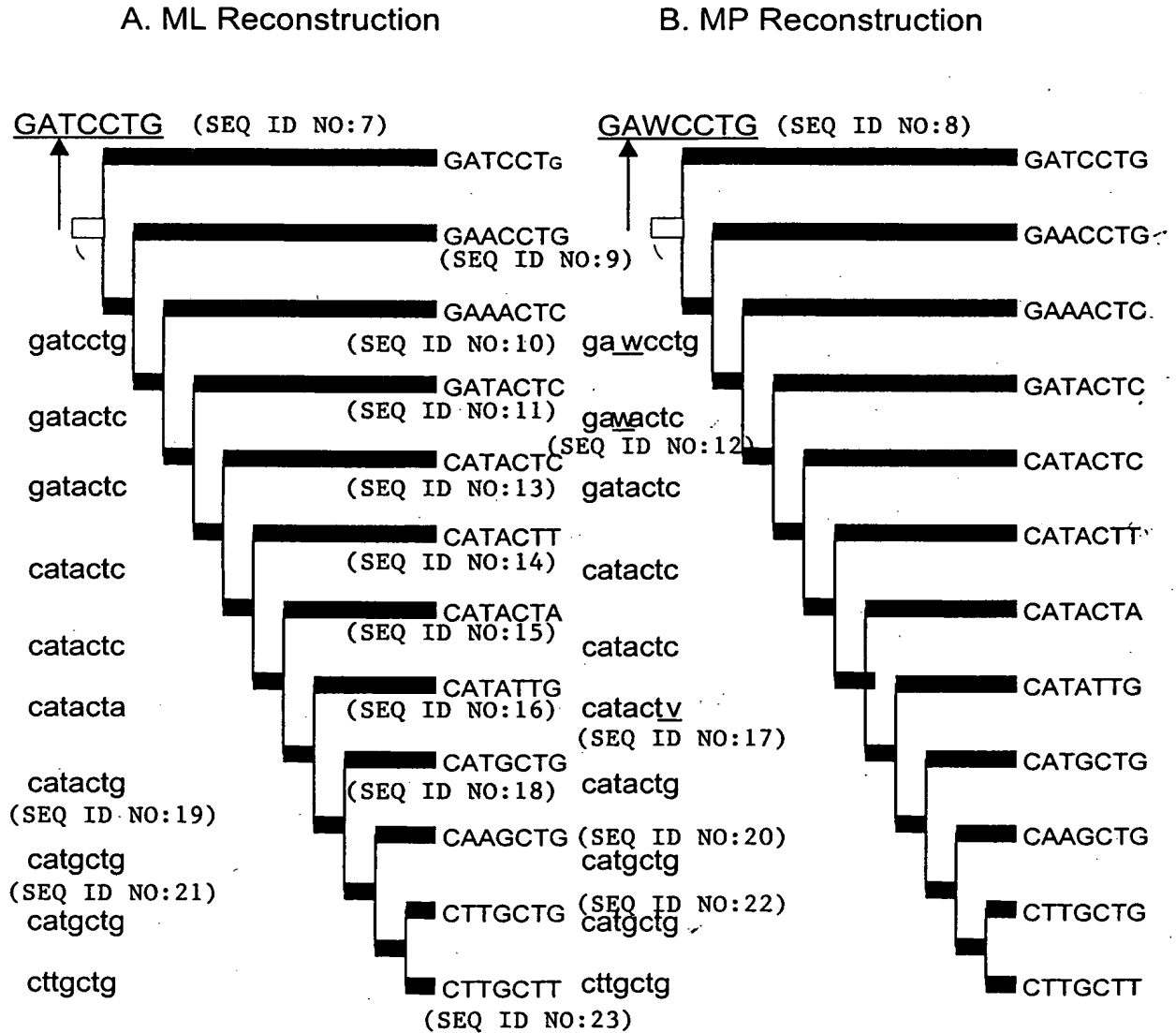
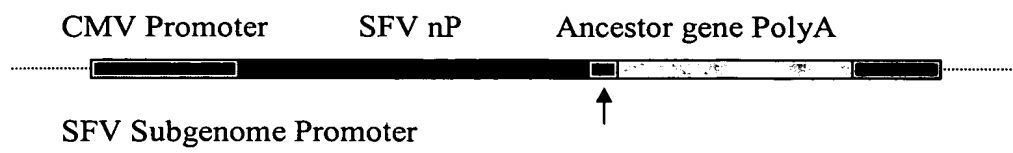


Figure 7



0.05 substitutions/site

Figure 8

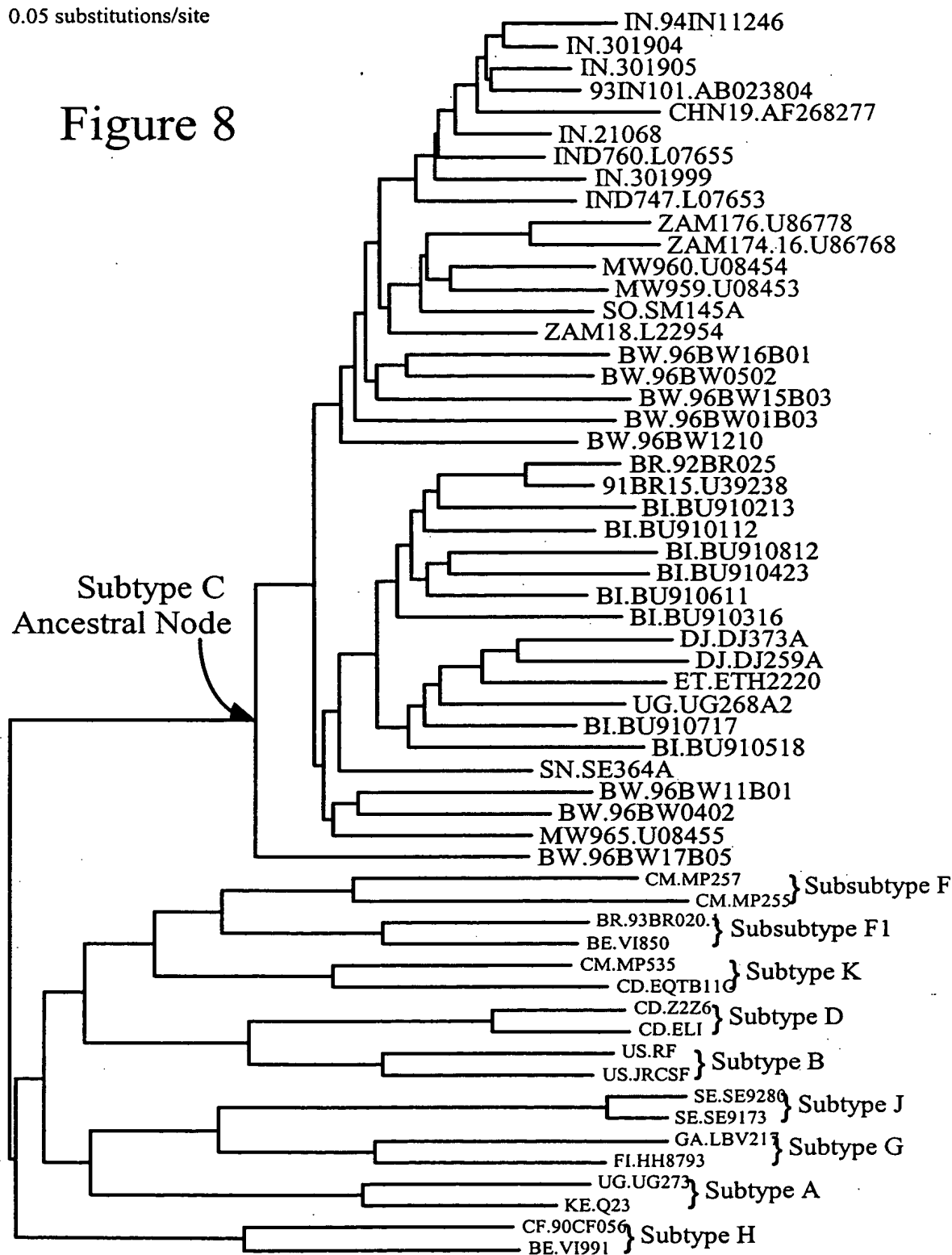


Figure 9
Comparison of Clade B gag Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Bgag.mrca	ATGGGTGCGA	GAGCGTCAGT	ATTAAGCGGG	GGAGAATTAG	ATAAATGGGA	AAAAATTCGG	TTACGGCCAG	GGGGAAAGAA	[80]
Bgag.LScotGGAGAAAA	[80]
Bgag.MMcotGGAGGAAAA	[80]
[90	100	110	120	130	140	150	160]	
[
Bgag.mrca	AAAATATAAA	TTAAACATA	TAGTATGGGC	AAGCAGGGAG	CTAGAACGAT	TCGCAGTTAA	TCCTGGCCTT	TTAGAAACAT	[160]
Bgag.LScotGGGGGGGG	[160]
Bgag.MMcotGGGGTCGG	[160]
[170	180	190	200	210	220	230	240]	
[
Bgag.mrca	CAGAAGGCTG	TAGACAAATA	CTGGGACAGC	TACAACCATC	CCTTCAGACA	GGATCAGAAG	AACTTAGATC	ATTATATAAT	[240]
Bgag.LScotCGATAAAA	[240]
Bgag.MMcotCGATAAAA	[240]
[250	260	270	280	290	300	310	320]	
[
Bgag.mrca	ACAGTAGCAG	TCCTCTATTG	TGTGCATCAA	AAGATAGAGG	TAAAAGACAC	CAAGGAAGCT	TTAGATAAGA	TAGAGGAAGA	[320]
Bgag.LScotACGGGTAA	[320]
Bgag.MMcotACGGTAAA	[320]
[330	340	350	360	370	380	390	400]	
[
Bgag.mrca	GCAAAACAAA	AGTAAGAAAA	AGGCACAGCA	AGCAGCAGCT	GACACAGGAA	ACAGCAGCCA	GGTCAGCCAA	AATTACCCTA	[400]
Bgag.LScotATGGACAC	[400]
Bgag.MMcotATGGACAC	[400]
[410	420	430	440	450	460	470	480]	
[
Bgag.mrca	TAGTGCAGAA	CCTACAGGGG	CAAATGGTAC	ATCAGGCCCT	ATCACCTAGA	ACTTTAAATG	CATGGGTAAA	AGTAATAGAA	[480]
Bgag.LScotCCAAAAGG	[480]
Bgag.MMcotTAGAAAAGG	[480]
[490	500	510	520	530	540	550	560]	
[
Bgag.mrca	GAGAAGGCTT	TCAGCCCAGA	AGTAATACCC	ATGTTTTCAG	CATTATCAGA	AGGAGCCACC	CCACAAGATT	TAAACACCAT	[560]
Bgag.LScotCCCCCCCC	[560]
Bgag.MMcotCCCCCCCC	[560]
[570	580	590	600	610	620	630	640]	
[
Bgag.mrca	GCTAAACACA	GTGGGGGGAC	ATCAAGCAGC	CATGCAAATG	TTAAAAGAGA	CCATCAATGA	GGAAGCTGCA	GAATGGGATA	[640]
Bgag.LScotAAAAAAAA	[640]
Bgag.MMcotAAAAAAAA	[640]
[650	660	670	680	690	700	710	720]	
[
Bgag.mrca	GATTGCATCC	AGTGCATGCA	GGGCCTATTG	CACCAGGCCA	GATGAGAGAA	CCAAGGGGAA	GTGACATAGC	AGGAACCTACT	[720]
Bgag.LScotCCCCCCCC	[720]
Bgag.MMcotCCCCCCCC	[720]
[730	740	750	760	770	780	790	800]	
[
Bgag.mrca	AGTACCCTTC	AGGAACAAAT	AGCATGGATG	ACAAATAATC	CACCTATCCC	AGTAGGAGAA	ATCTATAAAA	GATGGATAAT	[800]
Bgag.LScotGGGGGGGG	[800]
Bgag.MMcotGGGCGGGG	[800]
[810	820	830	840	850	860	870	880]	
[
Bgag.mrca	CCTGGGATTA	AATAAAATAG	TAAGAATGTA	TAGCCCTGTC	AGCATTTCTGG	ACATAAGACA	AGGACCAAAG	GAACCCTTTA	[880]
Bgag.LScotAAAACACACACAC	[880]
Bgag.MMcotAAAACACACACAC	[880]

[890	900	910	920	930	940	950	960]	
[
Bgag.mrca	GAGACTATGT	AGACCGGTTT	TATAAACTC	TAAGAGCCGA	GCAAGCTTCA	CAGGAGGTAA	AAAATTGGAT	GACAGAAACC	[960]
Bgag.LScot	[960]
Bgag.MMcotT....	T.....	[960]
[970	980	990	1000	1010	1020	1030	1040]	
[
Bgag.mrca	TTGTTGGTCC	AAAATGCGAA	CCCAGATTGT	AAGACTATCT	TAAAAGCATT	GGGACCAGGA	GCTACACTAG	AAGAAATGAT	[1040]
Bgag.LScotT..C..	[1040]
Bgag.MMcotC..T.	A.....	C.....	[1040]
[1050	1060	1070	1080	1090	1100	1110	1120]	
[
Bgag.mrca	GACAGCATGT	CAGGGAGTGG	GGGGACCCGG	CCATAAGCA	AGAGTTTGG	CTGAAGCAAT	GAGCCAAGTA	ACAAATTCAG	[1120]
Bgag.LScot	[1120]
Bgag.MMcotA..G...A.A.....C..	[1120]
[1130	1140	1150	1160	1170	1180	1190	1200]	
[
Bgag.mrca	CTACCATAAT	GATGCAGAGA	GGCAATTTTA	GGAACCCAAG	AAAGACTGTT	AAGTGTTCAT	ATTGTGGCAA	AGAAGGGCAC	[1200]
Bgag.LScotA...	[1200]
Bgag.MMcotG.....A.GG..A...A.....	[1200]
[1210	1220	1230	1240	1250	1260	1270	1280]	
[
Bgag.mrca	ATAGCCAGAA	ATTGCAGGGC	CCCTAGGAAA	AAGGGCTGTT	GGAAATGTGG	AAAGGAAGGA	CACCAAATGA	AAGATTGTAC	[1280]
Bgag.LScotA..	[1280]
Bgag.MMcotA.....	[1280]
[1290	1300	1310	1320	1330	1340	1350	1360]	
[
Bgag.mrca	TGAGAGACAG	GCTAATTTTT	TAGGGAAAAT	CTGGCCTTCC	CACAAGGGAA	GGCCAGGGAA	TTTCTTCAG	AGCAGACCAG	[1360]
Bgag.LScotG..	[1360]
Bgag.MMcotG..C..AG....	[1360]
[1370	1380	1390	1400	1410	1420	1430	1440]	
[
Bgag.mrca	AGCCAACAGC	CCCACCAGAA	GAGAGCTTCA	GGTTTGGGGA	AGAGACAACA	ACTCCCTCTC	AGAAGCAGGA	GCAGAAAGAC	[1440]
Bgag.LScot	[1440]
Bgag.MMcotG.....C..C..GG...	[1440]
[1450	1460	1470	1480	1490	1500]			
[
Bgag.mrca	AAGGAACTGT	ATCCTTTAGC	TTCCCTCAAA	TCACTCTTTG	GCAACGACCC	CTCGTCACAA	TAA	[1503]	
Bgag.LScotG..	[1503]	
Bgag.MMcotA..C..GA.G..A.....	[1503]	

Figure 10
Comparison of Clade B env Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Bgp160.mrca	ATGAGAGTGA	AGGGGATCAG	GAAGAATTGT	CAGCACTTGT	GGAAATGGGG	CACCATGCTC	CTTGGGATGT	TGATGATCTG	[80]
Bgp160.LScotA.G.	[80]
Bgp160.MMcotA.G.	[80]
[90	100	110	120	130	140	150	160]	
[
Bgp160.mrca	TAGTGCTGCA	GAAAACTTGT	GGGTCACAGT	CTATTATGGG	GTACCTGTGT	GGAAAGAAGC	AACCACCACT	CTATTTTGTG	[160]
Bgp160.LScotA.	[160]
Bgp160.MMcotA.	[160]
[170	180	190	200	210	220	230	240]	
[
Bgp160.mrca	CATCAGATGC	TAAAGCATAT	AAAACAGAGG	TACATAATGT	CTGGGCCACA	CATGCCTGTG	TACCCACAGA	CCCCAACCCA	[240]
Bgp160.LScotG.T.T.	[240]
Bgp160.MMcotG.T.T.	[240]
[250	260	270	280	290	300	310	320]	
[
Bgp160.mrca	CAAGAAGTAG	TATTGGAAAA	TGTGACAGAA	AATTTTAACA	TGTGGAAAAA	TAACATGGTA	GAACAGATGC	ATGAGGATAT	[320]
Bgp160.LScot	[320]
Bgp160.MMcot	[320]
[330	340	350	360	370	380	390	400]	
[
Bgp160.mrca	AATCAGTTTA	TGGGATCAAA	GCCTAAAGCC	ATGTGTAAAA	TTAACCCAC	TCTGTGTTAC	TTTAAATTGC	ACTGATGCGA	[400]
Bgp160.LScotTT.	[400]
Bgp160.MMcotTT.	[400]
[410	420	430	440	450	460	470	480]	
[
Bgp160.mrca	ACAAGAATGC	TACTAATACC	AATAGTAGTA	GTGGGGGAAC	AATGGAGAAA	GGAGAAATGA	AAAAGTGTCT	TTTCAATATC	[480]
Bgp160.LScot	.T.....C...A..T	G.....A.	[480]
Bgp160.MMcot	.T.....C...A..T	G.....A.	[480]
[490	500	510	520	530	540	550	560]	
[
Bgp160.mrca	ACCACAAGCA	TAAGAGATAA	GATGCAGAAA	GAATATGCAC	TTTTTTATAA	ACTTGATGTA	GTACCAATAG	ATAATGATAA	[560]
Bgp160.LScotG.	[560]
Bgp160.MMcotG.	[560]
[570	580	590	600	610	620	630	640]	
[
Bgp160.mrca	TAATAGTAAT	AATAATACCA	ACTATAGGTT	GATAAATTGT	AATACCTCAG	TCATTACACA	GGCCTGTCCA	AAGGTATCCT	[640]
Bgp160.LScotC.....C.....	G.....G.....	..C.....	[640]
Bgp160.MMcotC.....C.....	G.....G.....	..C.....	[640]
[650	660	670	680	690	700	710	720]	
[
Bgp160.mrca	TTGAGCCAAT	TCCCATACAT	TATTGTACCC	CGGCTGGTTT	TGCGATTCTA	AAGTGTAATG	ATAAGAAGTT	CAATGGAACA	[720]
Bgp160.LScotG...	[720]
Bgp160.MMcotG...	[720]
[730	740	750	760	770	780	790	800]	
[
Bgp160.mrca	GGACCATGTA	AAAATGTCAG	CACAGTACAA	TGTACACATG	GAATTAGGCC	AGTAGTGTCA	ACTCAACTGC	TGTTAAATGG	[800]
Bgp160.LScot	C.....A...	[800]
Bgp160.MMcot	C.....A...	[800]
[810	820	830	840	850	860	870	880]	
[
Bgp160.mrca	CAGTCTAGCA	GAAGAAGAGG	TAGTAATTAG	ATCTGAAAAT	TTCACGGACA	ATGCTAAAAC	CATAATAGTA	CAGCTGAATG	[880]
Bgp160.LScotC...	[880]
Bgp160.MMcotC...	[880]

[890	900	910	920	930	940	950	960]	
[
Bgp160.mrca	AATCTGTAGA	AATTAATTGT	ACAAGACCCA	ACAACAATAC	AAGAAAAAGT	ATACCTATAG	GACCAGGGAG	AGCACTTTAT	[960]
Bgp160.LScotA.....T.....	[960]
Bgp160.MMcotA.....T.....	[960]
[970	980	990	1000	1010	1020	1030	1040]	
[
Bgp160.mrca	ACAACAGGAG	AAATAATAGG	AGATATAAGA	CAAGCACATT	GTAACATTAG	TAGAGCAAAA	TGGAATAACA	CTTTAAAACA	[1040]
Bgp160.LScot	[1040]
Bgp160.MMcot	[1040]
[1050	1060	1070	1080	1090	1100	1110	1120]	
[
Bgp160.mrca	GGTAGTTACA	AAATTAAGAG	AACAATTTGG	GAATAATAAA	ACAATAGTCT	TTAATCCATC	CTCAGGAGGG	GACCCAGAAA	[1120]
Bgp160.LScot	.A.....A.A...	[1120]
Bgp160.MMcot	.A.....A.A...	[1120]
[1130	1140	1150	1160	1170	1180	1190	1200]	
[
Bgp160.mrca	TTGTAATGCA	CAGTTTTAAT	TGTGGAGGGG	AATTTTCTA	CTGTAATACA	ACACAACTGT	TTAATAGTAC	TTGGAATAGT	[1200]
Bgp160.LScotT..G..	[1200]
Bgp160.MMcotT..G..	[1200]
[1210	1220	1230	1240	1250	1260	1270	1280]	
[
Bgp160.mrca	ACTGAAGGGT	CAAATAAAAC	TACAGGGTCA	AATAACACTG	GAGGAGAAAC	TATCACACTC	CCATGCAGAA	TAAACAAAT	[1280]
Bgp160.LScot	...TGGACT.	GG....CT..	.GA.....	...G.....	A.....C..	[1280]
Bgp160.MMcot	...TGGACT.	GG....CT..	.GA.....	...G.....	A.....C..	[1280]
[1290	1300	1310	1320	1330	1340	1350	1360]	
[
Bgp160.mrca	TATAAACATG	TGGCAGGAAG	TAGGAAAAGC	AATGTATGCC	CCTCCCATCA	GAGGACAAAT	TAAATGTTCA	TCAAATATTA	[1360]
Bgp160.LScotG.....	[1360]
Bgp160.MMcotG.....	[1360]
[1370	1380	1390	1400	1410	1420	1430	1440]	
[
Bgp160.mrca	CAGGGCTACT	ATTAACAAGA	GATGGTGGTG	AAAATAGTAC	CAATGAGACC	GAGATCTTCA	GACCTGGAGG	AGGAGATATG	[1440]
Bgp160.LScotG..A	.T...AC..	...C.....	[1440]
Bgp160.MMcotG..A	.T...AC..	...C.....	[1440]
[1450	1460	1470	1480	1490	1500	1510	1520]	
[
Bgp160.mrca	AGGGACAATT	GGAGAAGTGA	ATTATATAAA	TATAAAGTAG	TAAAAATTGA	ACCATTAGGA	GTAGCACCCA	CCAAGGCAAA	[1520]
Bgp160.LScot	[1520]
Bgp160.MMcot	[1520]
[1530	1540	1550	1560	1570	1580	1590	1600]	
[
Bgp160.mrca	GAGAAGAGTG	GTGCAAAGAG	AAAAAAGAGC	AGTGGGAATA	ATAGGAGCTA	TGTTCTTGG	GTTCTTGGGA	GCAGCAGGAA	[1600]
Bgp160.LScotG...G	[1600]
Bgp160.MMcotG...G	[1600]
[1610	1620	1630	1640	1650	1660	1670	1680]	
[
Bgp160.mrca	GCACTATGGG	CGCAGCGTCA	ATGACGCTGA	CGGTACAGGC	CAGACAATTA	TTGTCTGGTA	TAGTGCAACA	GCAAAACAAT	[1680]
Bgp160.LScotG.....	[1680]
Bgp160.MMcotG.....	[1680]
[1690	1700	1710	1720	1730	1740	1750	1760]	
[
Bgp160.mrca	TTGCTGAGGG	CTATTGAGGC	GCAACAGCAT	CTGTTGCAAC	TCACGGTCTG	GGGCATCAAA	CAGCTCCAGG	CAAGAGTCCT	[1760]
Bgp160.LScotA.....G	[1760]
Bgp160.MMcotA.....G	[1760]

[1770	1780	1790	1800	1810	1820	1830	1840]	
[
Bgp160.mrca	GGCTGTGGAA	AGATACCTAA	GGGATCAACA	GCTCCTAGGA	ATTTGGGGTT	GCTCTGGA	AACTCATTTGC	ACCACTACTG	[1840]
Bgp160.LScotG..GG...	[1840]
Bgp160.MMcotG..GG...	[1840]
[1850	1860	1870	1880	1890	1900	1910	1920]	
[
Bgp160.mrca	TGCCTTGGAA	TGCTAGTTGG	AGTAATAAAT	CTCTGGATAA	GATTTGGAAT	AACATGACCT	GGATGGAGTG	GGAAAGAGAA	[1920]
Bgp160.LScotG.	[1920]
Bgp160.MMcotG.	[1920]
[1930	1940	1950	1960	1970	1980	1990	2000]	
[
Bgp160.mrca	ATTGACAATT	ACACAGGCTT	AATATACAAC	TTAATTGAAG	AATCGCAGAA	CCAGCAAGAA	AAGAATGAAC	AAGAATTATT	[2000]
Bgp160.LScotA....C....A..	...A.....	[2000]
Bgp160.MMcotA....C....A..	...A.....	[2000]
[2010	2020	2030	2040	2050	2060	2070	2080]	
[
Bgp160.mrca	GGAATTGGAT	AAGTGGGCAA	GTTTGTGGAA	TTGGTTTGAC	ATAACACAAT	GGCTGTGGTA	TATAAAAATA	TTCATAATGA	[2080]
Bgp160.LScotA...	..A.....A.C.	[2080]
Bgp160.MMcotA...	..A.....A.C.	[2080]
[2090	2100	2110	2120	2130	2140	2150	2160]	
[
Bgp160.mrca	TAGTAGGAGG	CTTGGTAGGT	TTAAGAATAG	TTTTTGCTGT	GCTTCTCTATA	GTGAATAGAG	TTAGGCAGGG	ATACTCACCA	[2160]
Bgp160.LScot	A.....	[2160]
Bgp160.MMcot	A.....	[2160]
[2170	2180	2190	2200	2210	2220	2230	2240]	
[
Bgp160.mrca	TTATCATTTT	AGACCCGCCT	CCCAGCCCCG	AGGGGACCCG	ACAGGCCCGA	AGGAATCGAA	GAAGAAGGTG	GAGAGAGAGA	[2240]
Bgp160.LScotG....	[2240]
Bgp160.MMcotG....	[2240]
[2250	2260	2270	2280	2290	2300	2310	2320]	
[
Bgp160.mrca	CAGAGACAGA	TCCGGTCGAT	TAGTGAATGG	ATTCTTAGCA	CTTATCTGGG	ACGATCTGCG	GAGCCTGTGC	CTCTTCAGCT	[2320]
Bgp160.LScotC..C....	[2320]
Bgp160.MMcotC..C....	[2320]
[2330	2340	2350	2360	2370	2380	2390	2400]	
[
Bgp160.mrca	ACCACCGCTT	GAGAGACTTA	CTCTTGATTG	TAGCGAGGAT	TGTGGAACCT	CTGGGACGCA	GGGGGTGGGA	AGCCCTCAAA	[2400]
Bgp160.LScotA.....	[2400]
Bgp160.MMcotA.....	[2400]
[2410	2420	2430	2440	2450	2460	2470	2480]	
[
Bgp160.mrca	TATTGGTGGA	ATCTCCTGCA	GTATTGGAGT	CAGGAATAA	AGAATAGTGC	TGTTAGCTTG	CTTAATGCCA	CAGCAATAGC	[2480]
Bgp160.LScotA..C.....C....	[2480]
Bgp160.MMcotA..C.....C....	[2480]
[2490	2500	2510	2520	2530	2540	2550	2560]	
[
Bgp160.mrca	AGTAGCTGAG	GGGACAGATA	GGGTATAGA	AGTAGTACAA	AGAGCTTGTA	GAGCTATTCT	TCACATACCT	AGAAGAATAA	[2560]
Bgp160.LScot	C.....	..C.....	[2560]
Bgp160.MMcot	C.....	..C.....	[2560]
[2570	2580							
[
Bgp160.mrca	GACAGGGCTT	AGAAAGGGCT	TTGCTATAA						[2589]
Bgp160.LScot	G.....						[2589]
Bgp160.MMcot	G.....						[2589]

Figure 11
Comparison of Clade B nef Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Bnef.mrca	ATGGGTGGCA	AGTGGTCAAA	ACGTAGTGTG	GTTGGATGGC	CTGCTGTAAG	GGAAAGAATG	AGACGAGCTG	AGCCAGCAGC	[80]
Bnef.MMcot	[80]
Bnef.LScot	[80]
[90	100	110	120	130	140	150	160]	
[
Bnef.mrca	AGATGGGGTG	GGAGCAGTAT	CTCGAGACCT	GGAAAAACAT	GGAGCAATCA	CAAGTAGCAA	TACAGCAGCT	ACTAATGCTG	[160]
Bnef.MMcot	[160]
Bnef.LScot	[160]
[170	180	190	200	210	220	230	240]	
[
Bnef.mrca	CTTGTGCCTG	GCTAGAAGCA	CAAGAGGAGG	AGGAGGTGGG	TTTTCCAGTC	AGACCTCAGG	TACCTTTAAG	ACCAATGACT	[240]
Bnef.MMcot	A.....	[240]
Bnef.LScot	A.....	[240]
[250	260	270	280	290	300	310	320]	
[
Bnef.mrca	TACAAGGCAG	CTGTAGATCT	TAGCCACTTT	TAAAAAGAAA	AGGGGGGACT	GGAAGGGCTA	GTTTACTCCC	AAAAAAGACA	[320]
Bnef.MMcotT.....	A.....	[320]
Bnef.LScotT.....	A.....	[320]
[330	340	350	360	370	380	390	400]	
[
Bnef.mrca	AGATATCCTT	GATCTGTGGG	TCTACCACAC	ACAAGGCTAC	TTCCCTGATT	GGCAGAACTA	CACACCAGGG	CCAGGGACCA	[400]
Bnef.MMcotT..	[400]
Bnef.LScotT..	[400]
[410	420	430	440	450	460	470	480]	
[
Bnef.mrca	GATATCCACT	GACCTTTGGA	TGGTGCTTCA	AGCTAGTACC	AGTTGAGCCA	GAGAAGGTAG	AAGAGGCCAC	TGAAGGAGAG	[480]
Bnef.MMcot	A.....	[480]
Bnef.LScot	A.....	[480]
[490	500	510	520	530	540	550	560]	
[
Bnef.mrca	AACAACAGCT	TGTTACACCC	TATGAGCCTG	CATGGAATGG	ATGACCCGGA	GAGAGAAGTG	TTAGTGTTGA	GGTTTGACAG	[560]
Bnef.MMcotT...A.G....A.....	A.....	[560]
Bnef.LScotG....A.....	A.....	[560]
[570	580	590	600	610	620]			
[
Bnef.mrca	CCGCCTAGCA	TTTCATCACA	TGGCCCGAGA	GAAGCATCCG	GAGTACTACA	AGGACTGCTG	A		[621]
Bnef.MMcotCT.....			[621]
Bnef.LScotCT.....			[621]

[890	900	910	920	930	940	950	960]	
[.]	
Bpol.mrca	CACCAGGGAT	TAGATATCAG	TACAATGTGC	TTCCACAGGG	ATGGAAAGGA	TCACCAGCAA	TATTCCAAAG	TAGCATGACA	[960]
Bpol.LScot	[960]
Bpol.MMcot	[960]
[970	980	990	1000	1010	1020	1030	1040]	
[.]	
Bpol.mrca	AAAATCTTAG	AGCCTTTTAG	AAAACAAAAT	CCAGAAATAG	TTATCTATCA	ATACATGGAT	GATTTGTATG	TAGGATCTGA	[1040]
Bpol.LScotC.....	[1040]
Bpol.MMcotC.....	[1040]
[1050	1060	1070	1080	1090	1100	1110	1120]	
[.]	
Bpol.mrca	CTTAGAAATA	GGGCAGCATA	GAACAAAAAT	AGAGGAACTG	AGAGAACATC	TGTTGAGGTG	GGGATTTACC	ACACCAGACA	[1120]
Bpol.LScotC.....	[1120]
Bpol.MMcotC.....	[1120]
[1130	1140	1150	1160	1170	1180	1190	1200]	
[.]	
Bpol.mrca	AAAAACATCA	GAAAGAACCT	CCATTCTTTT	GGATGGGTTA	TGAACTCCAT	CCTGATAAAT	GGACAGTACA	GCCTATAGTG	[1200]
Bpol.LScotC.....	[1200]
Bpol.MMcotC.....	[1200]
[1210	1220	1230	1240	1250	1260	1270	1280]	
[.]	
Bpol.mrca	CTGCCAGAAA	AAGACAGCTG	GACTGTCAAT	GACATACAGA	AGTTAGTGGG	AAAATTGAAT	TGGGCAAGTC	AGATTTATGC	[1280]
Bpol.LScot	[1280]
Bpol.MMcotCC.	[1280]
[1290	1300	1310	1320	1330	1340	1350	1360]	
[.]	
Bpol.mrca	AGGGATTAAA	GTAAGCAAT	TATGTAAACT	CCTTAGGGGA	ACCAAAGCAC	TAACAGAAGT	AGTACCACTA	ACAGAAGAAG	[1360]
Bpol.LScotA.....	[1360]
Bpol.MMcotA.....	[1360]
[1370	1380	1390	1400	1410	1420	1430	1440]	
[.]	
Bpol.mrca	CAGAGCTAGA	ACTGGCAGAA	AACAGGGAGA	TTCTAAAAGA	ACCAGTACAT	GGAGTGTATT	ATGACCCATC	AAAAGACTTA	[1440]
Bpol.LScot	[1440]
Bpol.MMcotA.....	[1440]
[1450	1460	1470	1480	1490	1500	1510	1520]	
[.]	
Bpol.mrca	ATAGCAGAAA	TACAGAAGCA	GGGGCAAGGC	CAATGGACAT	ATCAAATTTA	TCAAGAGCCA	TTTAAAAATC	TGAAAACAGG	[1520]
Bpol.LScot	[1520]
Bpol.MMcot	[1520]
[1530	1540	1550	1560	1570	1580	1590	1600]	
[.]	
Bpol.mrca	AAAGTATGCA	AGAATGAGGG	GTGCCCACAC	TAATGATGTA	AAACAATTAA	CAGAGGCAGT	GCAAAAAATA	GCCACAGAAA	[1600]
Bpol.LScot	[1600]
Bpol.MMcot	...A.....	[1600]
[1610	1620	1630	1640	1650	1660	1670	1680]	
[.]	
Bpol.mrca	GCATAGTAAT	ATGGGGAAAG	ACTCCTAAAT	TTAACTACC	CATACAAAAG	GAAACATGGG	AAGCATGGTG	GACAGAGTAT	[1680]
Bpol.LScotA.....	[1680]
Bpol.MMcotA.....A.....	[1680]
[1690	1700	1710	1720	1730	1740	1750	1760]	
[.]	
Bpol.mrca	TGGCAAGCCA	CCTGGATTCC	TGAGTGGGAG	TTTGTCAATA	CCCCTCCCTT	AGTAAAAATTA	TGGTACCAGT	TAGAGAAAGA	[1760]
Bpol.LScotG.....	[1760]
Bpol.MMcotG.....	[1760]
[1770	1780	1790	1800	1810	1820	1830	1840]	
[.]	
Bpol.mrca	ACCCATAGTA	GGAGCAGAAA	CTTTCTATGT	AGATGGGGCA	GCTAATAGAG	AGACTAAATT	AGGAAAAGCA	GGATATGTTA	[1840]
Bpol.LScotG.....	[1840]
Bpol.MMcotC.....G.....	[1840]

[1850	1860	1870	1880	1890	1900	1910	1920]	
[
Bpol.mrca	CTGACAGAGG	AAGACAAAAA	GTTGTCTCCC	TAAGTGACAC	AACAAATCAG	AAGACTGAGT	TACAAGCAAT	TCATCTAGCT		[1920]
Bpol.LScot		[1920]
Bpol.MMcot	..A.....		[1920]
[1930	1940	1950	1960	1970	1980	1990	2000]	
[
Bpol.mrca	TTGCAGGATT	CGGGATTAGA	AGTAAACATA	GTAACAGACT	CACAAATATGC	ATTAGGAATC	ATTCAAGCAC	AACCAGATAA		[2000]
Bpol.LScot		[2000]
Bpol.MMcot		[2000]
[2010	2020	2030	2040	2050	2060	2070	2080]	
[
Bpol.mrca	GAGTGAATCA	GAGTTAGTCA	GTCAAATAAT	AGAGCAGTTA	ATAAAAAAGG	AAAAGGTCTA	CCTGGCATGG	GTACCAGCAC		[2080]
Bpol.LScot		[2080]
Bpol.MMcot	A.....		[2080]
[2090	2100	2110	2120	2130	2140	2150	2160]	
[
Bpol.mrca	ACAAAGGAAT	TGGAGGAAAT	GAACAAGTAG	ATAAATTAGT	CAGTACTGGA	ATCAGGAAAG	TACTATTTTT	GGATGGAATA		[2160]
Bpol.LScotG.....		[2160]
Bpol.MMcotG.....	A.....		[2160]
[2170	2180	2190	2200	2210	2220	2230	2240]	
[
Bpol.mrca	GATAAGGCCC	AAGAAGAACA	TGAGAAATAT	CACAGTAATT	GGAGAGCAAT	GGCTAGTGAT	TTTAACCTGC	CACCTGTAGT		[2240]
Bpol.LScot		[2240]
Bpol.MMcot		[2240]
[2250	2260	2270	2280	2290	2300	2310	2320]	
[
Bpol.mrca	AGCAAAAGAA	ATAGTAGCCA	GCTGTGATAA	ATGTCAGCTA	AAAGGAGAAG	CCATGCATGG	ACAAGTAGAC	TGTAGTCCAG		[2320]
Bpol.LScot		[2320]
Bpol.MMcot		[2320]
[2330	2340	2350	2360	2370	2380	2390	2400]	
[
Bpol.mrca	GAATATGGCA	ACTAGATTGT	ACACATTTAG	AAGGAAAAGT	TATCCTGGTA	GCAGTTCATG	TAGCCAGTGG	CTATATAGAA		[2400]
Bpol.LScot	A.....		[2400]
Bpol.MMcot	A.....		[2400]
[2410	2420	2430	2440	2450	2460	2470	2480]	
[
Bpol.mrca	GCAGAAGTTA	TTCCAGCAGA	AACAGGGCAG	GAAACAGCAT	ACTTTCTCTT	AAAATTAGCA	GGAAGATGGC	CAGTAAAAGT		[2480]
Bpol.LScot	G.....AC		[2480]
Bpol.MMcot	G.....AC		[2480]
[2490	2500	2510	2520	2530	2540	2550	2560]	
[
Bpol.mrca	AATACATACA	GACAATGGCA	GCAATTTTAC	CAGTACTACA	GTTAAGGCCG	CCTGTTGGTG	GGCAGGGATC	AAGCAGGAAT		[2560]
Bpol.LScotG.....		[2560]
Bpol.MMcotG.....G.....		[2560]
[2570	2580	2590	2600	2610	2620	2630	2640]	
[
Bpol.mrca	TTGGCATTCC	CTACAATCCC	CAAAGTCAAG	GAGTAGTAGA	ATCTATGAAT	AAAGAATTAA	AGAAAATTAT	AGGACAGGTA		[2640]
Bpol.LScot		[2640]
Bpol.MMcot		[2640]
[2650	2660	2670	2680	2690	2700	2710	27	

```

[          2730      2740      2750      2760      2770      2780      2790      2800]
[          .          .          .          .          .          .          .          .]
Bpol.mrca  GGGGTACAGT GCAGGGGAAA GAATAGTAGA CATAATAGCA ACAGACATAC AAAGTAAAGA ACTACAAAAA CAAATTACAA [2800]
Bpol.LScot .....          .....          .....          .....          .....          .T.....          ..... [2800]
Bpol.MMcot .....          .....          .....          .....          .....          .T.....          ..... [2800]

[          2810      2820      2830      2840      2850      2860      2870      2880]
[          .          .          .          .          .          .          .          .]
Bpol.mrca  AAATTCAAAA TTTTCGGGTT TATTACAGGG ACAGCAGAGA TCCACTTTGG AAAGGACCAG CAAAGCTTCT CTGGAAAGGT [2880]
Bpol.LScot .....          .....          .....          .....          .....          .....          ..... [2880]
Bpol.MMcot .....          .....          .....          .....          .....          .....          ..... [2880]

[          2890      2900      2910      2920      2930      2940      2950      2960]
[          .          .          .          .          .          .          .          .]
Bpol.mrca  GAAGGGGCAG TAGTAATACA AGATAATAGT GACATAAAAG TAGTGCCAAG AAGAAAAGCA AAGATCATTG GGGATTATGG [2960]
Bpol.LScot .....          .....          .....          .....          .....          .....          ..... [2960]
Bpol.MMcot .....          .....          .....          .....          .....          .....          ..... [2960]

[          2970      2980      2990      3000      3010]
[          .          .          .          .          .]
Bpol.mrca  AAAACAGATG GCAGGTGATG ATTGTGTGGC AAGTAGACAG GATGAGGATT AG [3012]
Bpol.LScot .....          .....          .....          .....          ..... [3012]
Bpol.MMcot .....          .....          .....          .....          ..... [3012]

```

Figure 13
Comparison of Clade B rev Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.	
Brev.mrca	ATGGCAGGAA	GAAGCGGAGA	CAGCGACGAA	GAGCTCCTCA	AGACAGTCAG	ACTCATCAAG	TTTCTCTATC	AAAGCAACCC	[80]
Brev.LScot	[80]
Brev.MMcot	[80]
[90	100	110	120	130	140	150	160]	
[.	
Brev.mrca	GCCTCCCAGC	CCCGAGGGGA	CCCGACAGGC	CCGAAGGAAT	AGAAGAAGAA	GGTGGAGAGA	GAGACAGAGA	CAGATCCGTT	[160]
Brev.LScot	C.....G.	[160]
Brev.MMcot	C.....G.	[160]
[170	180	190	200	210	220	230	240]	
[.	
Brev.mrca	CGATTAGTGA	ACGGATTCTT	AGCACTTATC	TGGGACGATC	TGCGGAGCCT	GTGCCTCTTC	AGCTACCACC	GCTTGAGAGA	[240]
Brev.LScotT.....T...C.	[240]
Brev.MMcotT.....C.	[240]
[250	260	270	280	290	300	310	320]	
[.	
Brev.mrca	CTTACTCTTG	ATTGTAGCGA	GGATTGTGGA	ACTTCTGGGA	CGCAGGGGGT	GGGAAGTCCT	CAAATATTGG	TGGAATCTCC	[320]
Brev.LScotA...	[320]
Brev.MMcot	[320]
[330	340	350	360]					
[.				.	
Brev.mrca	TGCACTATTG	GAGTCAGGAA	CTAAAGAATA	GTGCTGTTAG	[360]				
Brev.LScot	.A.....	[360]				
Brev.MMcot	[360]				

Figure 14
Comparison of Clade B tat Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.	
Btat.mrca	ATGGAGCCAG	TAGATCCTAG	ACTAGAGCCC	TGGAAGCATC	CAGGAAGTCA	GCCTAAGACT	GCTTGTACCA	ATTGCTATTG	[80]
Btat.LScot	[80]
Btat.MMcot	[80]
[90	100	110	120	130	140	150	160]	
[.	
Btat.mrca	TAAAAAGTGT	TGCTATCATT	GCCAAGTTTG	CTTCATAACA	AAAGGCTTAG	GCATCTCCTA	TGGCAGGAAG	AAGCGGAGAC	[160]
Btat.LScotT.....	T.....	[160]
Btat.MMcotT.....	T.....	[160]
[170	180	190	200	210	220	230	240]	
[.	
Btat.mrca	AGCGACGAAG	ACCTCCTCAA	GGCAGTCAGA	CTCATCAAGT	TTCTCTATCA	AAGCAACCCG	CCTCCCAGCC	CCGAGGGGAC	[240]
Btat.LScotG.....	.A.....	[240]
Btat.MMcotG.....	.A.....	[240]
[250	260	270	280	290	300	310	320]	
[.	
Btat.mrca	CCGACAGGCC	CGAAGGAATC	GAAGAAGAAG	GTGGAGAGAG	AGACAGAGAC	AGATCCGGTC	GATTAGTGAA	TGGATTCTTA	[320]
Btat.LScot	[320]
Btat.MMcotG.	[320]
[
[
Btat.mrca	G	[321]							
Btat.LScot	.	[321]							
Btat.MMcot	T	[321]							

Figure 15
Comparison of Clade B vif Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.]	
Bvif.mrca	ATGGA	AAACA	GATGG	CAGGT	GATGATT	GTG	TGGCA	AGTAG	ACAGGATGAG GATTAGAACA TGGAAAAGTT TAGTAAACA [80]
Bvif.LScot	[80]
Bvif.MMcot	[80]
[90	100	110	120	130	140	150	160]	
[.]	
Bvif.mrca	CCATATGTAT	ATTTCAAAGA	AAGCTAAGGG	ATGGTTT	TAT	AGACATCACT	ATGAAAGCAC	TCATCCAAGA	ATAAGTTCAG [160]
Bvif.LScotG	[160]
Bvif.MMcotG	[160]
[170	180	190	200	210	220	230	240]	
[.]	
Bvif.mrca	AAGTACACAT	CCCACTAGGA	GATGCTAGAT	TGGTAATAAA	AACATATTGG	GGTCTGCATA	CAGGAGAAAG	AGAATGGCAT	[240]
Bvif.LScotGCC	[240]
Bvif.MMcotGCC	[240]
[250	260	270	280	290	300	310	320]	
[.]	
Bvif.mrca	TTGGGTCAGG	GAGTCTCCAT	AGAATGGAGG	AAAAGGAGAT	ATAGCACACA	AGTAGACCCT	GGCCTAGCAG	ACCAACTAAT	[320]
Bvif.LScotAA	[320]
Bvif.MMcotAA	[320]
[330	340	350	360	370	380	390	400]	
[.]	
Bvif.mrca	TCATCTGTAT	TATTTTGATT	GTTTTTCAGA	ATCTGCTATA	AGAAATGCCA	TATTAGGACA	TATAGTTAGT	CCTAGGTGTG	[400]
Bvif.LScotC	[400]
Bvif.MMcotC	[400]
[410	420	430	440	450	460	470	480]	
[.]	
Bvif.mrca	AATATCAAGC	AGGACATAAC	AAGGTAGGAT	CTCTACAGTA	CTTGCCACTA	ACAGCATTAA	TAACACCAAA	AAAGATAAAG	[480]
Bvif.LScotG	[480]
Bvif.MMcotG	[480]
[490	500	510	520	530	540	550	560]	
[.]	
Bvif.mrca	CCACCTTTGC	CTAGTGTTAG	GAAACTGACA	GAGGATAGAT	GGAACAAGCC	CCAGAAGACC	AAGGGCCACA	GAGGGAGCCA	[560]
Bvif.LScotC	[560]
Bvif.MMcotC	[560]
[570]							
[.]							
Bvif.mrca	TACAATGAAT	GGACACTAG	[579]						
Bvif.LScot	[579]						
Bvif.MMcot	[579]						

Figure 16
Comparison of Clade B vpr Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Bvpr.mrca	ATGGAACAAG	CCCCAGAAGA	CCAAGGGCCA	CAGAGGGAGC	CATACAATGA	ATGGACACTA	GAGCTTTTAG	AGGAGCTTAA	[80]
Bvpr.LScot	[80]
Bvpr.MMcot	[80]
[90	100	110	120	130	140	150	160]	
[
Bvpr.mrca	GAGTGAAGCT	GTTAGACATT	TTCCTAGGCT	ATGGCTCCAT	AGCTTAGGAC	AACATATCTA	TGAAACTTAT	GGGGATACCT	[160]
Bvpr.LScot	A.	T.	[160]
Bvpr.MMcot	A.	T.	[160]
[170	180	190	200	210	220	230	240]	
[
Bvpr.mrca	GGGCAGGAGT	GGAAGCTATA	ATAAGAATTC	TGCAACAAC	GCTGTTTATT	CATTTTCAGAA	TTGGGTGTCA	ACATAGCAGA	[240]
Bvpr.LScot	C.	G	[240]
Bvpr.MMcot	C.	[240]
[250	260	270	280	290]				
[
Bvpr.mrca	ATAGGCATTA	CTCGACAGAG	AAGAGCAAGA	AATGGAGCCA	G	[291]
Bvpr.LScot	G.	[291]
Bvpr.MMcot	G.	[291]

Figure 17
Comparison of Clade B vpu Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Bvpu.mrca	ATGCAACCTT	TAGAAATATT	AGCAATAGTA	GCATTAGTAG	TAGCAGCAAT	ACTAGCAATA	GTGTGTGGA	CCATAGTATT	[80]
Bvpu.LScotC.....A.....	[80]
Bvpu.MMcotC.....A.....	[80]
[90	100	110	120	130	140	150	160]	
[
Bvpu.mrca	CATAGAATAT	AGGAAAATAT	TAAGGCAAAG	AAAAATAGAC	AGGTTAATTG	ATAGAATAAG	AGAAAGAGCA	GAAGACAGTG	[160]
Bvpu.LScotA.....	[160]
Bvpu.MMcotA.....	[160]
[170	180	190	200	210	220	230	240]	
[
Bvpu.mrca	GCAATGAGAG	TGAAGGGGAT	CAGGAAGAAT	TATCAGCACT	TGTGGAAATG	GGGCACCATG	CTCCTTGGGA	TGTTGATGAT	[240]
Bvpu.LScotG.....	[240]
Bvpu.MMcotG.....	[240]
[
[
Bvpu.mrca	CTGTAG	[246]							
Bvpu.LScot	[246]							
Bvpu.MMcot	[246]							

Figure 18
Comparison of Clade B gag Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Bgag.mrca	MGARASVLSG	GELDKWEKIR	LRPGGKKKYK	LKHIVWASRE	LERFAVNPGL	LETSEGCRQI	LGQLQPSLQT	GSEELRSLYN	[80]
Bgag.LScotR.....R.....R.....R.....R.....R.....R.....K.....	[80]
Bgag.MMcot	...G.....	...K...R.....R.....R.....R.....R.....	...E...H.....K.....	[80]
[90	100	110	120	130	140	150	160]	
[
Bgag.mrca	TVAVLYCVHQ	KIEVKDTKEA	LDKIEEEQNK	SKKKAQQAAA	DTGNSSQVSQ	NYPIVQNLQG	QMVHQALSPR	TLNAWVKVIE	[160]
Bgag.LScot	...T.....	R.....	...E.....R.....R.....R.....I.....V.....	[160]
Bgag.MMcot	...T.....	N...R...D.	...E.....	I...R.....NP...M.....I.....V.....	[160]
[170	180	190	200	210	220	230	240]	
[
Bgag.mrca	EKAFSPEVIP	MFSALSEGAT	PQDLNTMLNT	VGGHQAAMQM	LKETINEEAA	EWDRLHPVHA	GPIAPGQMRE	PRGSDIAGTT	[240]
Bgag.LScot	[240]
Bgag.MMcot	[240]
[250	260	270	280	290	300	310	320]	
[
Bgag.mrca	STLQEQIAWM	TNNPPIPVGE	IYKRWIILGL	NKIVRMYSVP	SILDIRQGPK	EPFRDYVDRF	YKTLRAEQAS	QEVKNWMTET	[320]
Bgag.LScotG..T.....	[320]
Bgag.MMcotG..	...H.....M.....T.....	[320]
[330	340	350	360	370	380	390	400]	
[
Bgag.mrca	LLVQNaNPDC	KTILKALGPG	ATLEEMMTAC	QGVGGPGHKA	RVLAEAMSQV	TNSATIMMQR	GNFRNPRTKV	KCFNCGKEGH	[400]
Bgag.LScotA.....Q.....	[400]
Bgag.MMcotA.....S.....A.....KGQ....	[400]
[410	420	430	440	450	460	470	480]	
[
Bgag.mrca	IARNCRAPRK	KGCWCKGKEG	HQMKDCTERQ	ANFLGKIWPS	HKGRPGNFLQ	SRPEPTAPPE	ESFRFGEETT	TPSQKQEQKD	[480]
Bgag.LScot	..K.....PI.	[480]
Bgag.MMcotP....PR.	[480]
[490	500]							
[
Bgag.mrca	KELYPLASLK	SLFGNDPSSQ	[500]						
Bgag.LScotR.....	[500]						
Bgag.MMcot	..Q...T..R	[500]						

Figure 19
Comparison of Clade B gp160 Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Bgp160.mrca	MRVKGIRKNC	QHLWKWGTML	LGMLMICSAA	ENLWVTVYVG	VPVWKEATT	LFCASDAKAY	KTEVHNVWAT	HACVPTDPNP	[80]
Bgp160.LScotYRKD	[80]
Bgp160.MMcotYRKD	[80]
[90	100	110	120	130	140	150	160]	
[
Bgp160.mrca	QEVVLENVTE	NFNMWKNNMV	EQMHEDIISL	WDQSLKPCVK	LTPLCVTLNC	TDANKNATNT	NSSSGGTMEK	GEMKNCSFNI	[160]
Bgp160.LScotLEMI	[160]
Bgp160.MMcotLEMI	[160]
[170	180	190	200	210	220	230	240]	
[
Bgp160.mrca	TTSIRDKMQK	EYALFYKLDV	VPIDNDNNSN	NNTNYRLINC	NTSVITQACP	KVSFEPIPIH	YCTPAGFAIL	KCNDKKFNGT	[240]
Bgp160.LScotVTT.SSA	[240]
Bgp160.MMcotVTT.SSA	[240]
[250	260	270	280	290	300	310	320]	
[
Bgp160.mrca	GPCKNVSTVQ	CTHGIRPVVS	TQLLLNGSLA	EEEVVIRSEN	FTDNAKTIIV	QLNESVEINC	TRPNNNTRKS	IPIGPGRALY	[320]
Bgp160.LScot	...TDH	[320]
Bgp160.MMcot	...TDH	[320]
[330	340	350	360	370	380	390	400]	
[
Bgp160.mrca	TTGEIIGDIR	QAHCNISRAK	WNNTLKQVV-	-TKLREQFGN	NKTIVFNPS	GGDPEIVMHS	FNCGGEFFYC	NTTQLFNSTW	[398]
Bgp160.LScotI	-KQS	[398]
Bgp160.MMcotI	-KQS	[398]
[410	420	430	440	450	460	470	480]	
[
Bgp160.mrca	NSTEGSNKTT	GSNNTGGETI	TLPCRIKQII	NMWQEVGKAM	YAPPIRGQIK	CSSNITGLLL	TRDGGENSTN	ETEIFRPGGG	[478]
Bgp160.LScot	.G.WTW.T.E	...D.E.DRN.N	[478]
Bgp160.MMcot	.G.WTW.T.E	...D.E.DRN.N	[478]
[490	500	510	520	530	540	550	560]	
[
Bgp160.mrca	DMRDNRSEL	YKYKVVKIEP	LGVAPTKAKR	RVVQREKRAV	GIIGAMFLGF	LGAAGSTMGA	ASMTLTVQAR	QLLSGIVQQQ	[558]
Bgp160.LScotV	[558]
Bgp160.MMcotV	[558]
[570	580	590	600	610	620	630	640]	
[
Bgp160.mrca	NNLLRAIEAQ	QHLLQLTVWG	IKQLQARVLA	VERYLRDQQL	LGIWGCSGKL	ICTTTVPWNA	SWSNKS LDKI	WNNMTWMEWE	[638]
Bgp160.LScotAE	[638]
Bgp160.MMcotAE	[638]
[650	660	670	680	690	700	710	720]	
[
Bgp160.mrca	REIDNYTGLI	YNLIEESQNG	QEKNEQELLE	LDKWASLWNW	FDITQWLWYI	KIFIMIVGGL	VGLRIVFAVL	SIVNRVRQGY	[718]
Bgp160.LScotSTN	[718]
Bgp160.MMcotSTN	[718]
[730	740	750	760	770	780	790	800]	
[
Bgp160.mrca	SPLSFQTRLR	APRGPDRPEG	IEEEGGERDR	DRSGRLVNGF	LALIWDLLRS	LCLFSYHRLR	DLLLIVARIV	ELLGRRGWEA	[798]
Bgp160.LScotT	[798]
Bgp160.MMcotT	[798]
[810	820	830	840	850	860]		
[
Bgp160.mrca	LKYWWNLLQY	WSQELKNSAV	SLLNATAIAV	AEGTDRVIEV	VQRACRAILH	IPRRIRQGLE	RALL		[862]
Bgp160.LScotT		[862]
Bgp160.MMcotT		[862]

Figure 20
Comparison of Clade B nef Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Bnef.mrca	MGGKWSKRSV	VGWPAVRERM	RRAEPAADGV	GAVSRDLEKH	GAITSSNTAA	TNAACAWLEA	QEEEEVGFPV	RPQVPLRPMT	[80]
Bnef.MMcotD.....	[80]
Bnef.LScotD.....	[80]
[90	100	110	120	130	140	150	160]	
[
Bnef.mrca	YKAAVDLSHF	LKEKGGLEGL	VYSQKRQDIL	DLWVYHTQGY	FPDWQNYTPG	PGTRYPLTFG	WCFKLVPVEP	EKVEEATEGE	[160]
Bnef.MMcotL.....I.....I.....N...	[160]
Bnef.LScotL.....I.....I.....N...	[160]
[170	180	190	200]				
[
Bnef.mrca	NNSLLHPMSL	HGMDDPEREV	LVVRFDSRLA	FHHMAREKHP	EYYKDC				[206]
Bnef.MMcot	..C.....QK..	...K.....L..				[206]
Bnef.LScotK..	...K.....L..				[206]

Figure 21
Comparison of Clade B pol Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Bpol.mrca	FFRENLAFFQ	GKARELSSEQ	TRANSPTRRE	LQVWGRDNNS	LSEAGADRQG	TVSFSFPQIT	LWQRPLVTIK	IGGQLKEALL	[80]
Bpol.LScotD....F....							[80]
Bpol.MMcotD....F....							[80]
[90	100	110	120	130	140	150	160]	
[
Bpol.mrca	DTGADDTVLE	EMNLPQKWKP	KMIGGIGGFI	KVRQYDQIPI	EICGHKAIGT	VLVGPTPVNI	IGRNLLTQIG	CTLNFPISPI	[160]
Bpol.LScotR....								[160]
Bpol.MMcotR....			L					[160]
[170	180	190	200	210	220	230	240]	
[
Bpol.mrca	ETVPVKLKPG	MDGPKVKQWP	LTEEKIKALV	EICTEMEKEG	KISKIGPENP	YNTPVFAIKK	KDSTKWRKLV	DFRELNKRTO	[240]
Bpol.LScot								[240]
Bpol.MMcot								[240]
[250	260	270	280	290	300	310	320]	
[
Bpol.mrca	DFWEVQLGIP	HPAGLKKKKS	VTVLVDVGDY	FSVPLDEDFR	KYTAFTIPSI	NNETPGIRYQ	YNVLPQGWKG	SPAIFQSSMT	[320]
Bpol.LScot								[320]
Bpol.MMcot								[320]
[330	340	350	360	370	380	390	400]	
[
Bpol.mrca	KILEPFRKQN	PEIVIIQYMD	DLYVGSdleI	GQHRTKIEEL	REHLLRWGFT	TPDKKHQKEP	PFLWMGYELH	PDKWTVQPIV	[400]
Bpol.LScotD....			Q					[400]
Bpol.MMcotD....			Q					[400]
[410	420	430	440	450	460	470	480]	
[
Bpol.mrca	LPEKDSWTVN	DIQKLVGKLN	WASQIYAGIK	VKQLCKLLRG	TKALTEVVPL	TEEAELAE	NREILKEPVH	GVYDPSKDL	[480]
Bpol.LScot				I				[480]
Bpol.MMcotP...				I				[480]
[490	500	510	520	530	540	550	560]	
[
Bpol.mrca	IAEIQKQGQG	QWTYQIYQEP	FKNLKTGKYA	RMRGAHTNDV	KQLTEAVQKI	ATESIVIWGK	TPKFKLPIQK	ETWEAWWTEY	[560]
Bpol.LScot								[560]
Bpol.MMcot						T		[560]
[570	580	590	600	610	620	630	640]	
[
Bpol.mrca	WQATWIPEWE	FVNTPLVLKL	WYQLEKEPIV	GAETFYVDGA	ANRETKLGKA	GYVTDGRQK	VVSLTDTTNQ	KTELQAIHLA	[640]
Bpol.LScot								[640]
Bpol.MMcot				N				[640]
[650	660	670	680	690	700	710	720]	
[
Bpol.mrca	LQDSGLEVNI	VTDSQYALGI	IQAQPKSES	ELVSQIIEQL	IKKEKVYLAW	VPAHKGIGGN	EQVDKLVSTG	IRKVLFLDGI	[720]
Bpol.LScot						A		[720]
Bpol.MMcot						A		[720]
[730	740	750	760	770	780	790	800]	
[
Bpol.mrca	DKAQEEHEKY	HSNWRAMASD	FNLPPVVAKE	IVASCDKCQL	KGEAMHGQVD	CSPGIWQLDC	THLEGKVLIV	AVHVASGYIE	[800]
Bpol.LScot								[800]
Bpol.MMcot								[800]
[810	820	830	840	850	860	870	880]	
[
Bpol.mrca	AEVIPAETGQ	ETAYFLKLKLA	GRWPVKVIHT	DNGSNFTSTT	VKAACWAGI	KQEFGIPYNP	QSQGVVESMN	KELKKIIGQV	[880]
Bpol.LScot		T						[880]
Bpol.MMcot		T						[880]

```

[      890      900      910      920      930      940      950      960]
[      .      .      .      .      .      .      .      .]
Bpol.mrca  RDQAEHLKTA VQMAVFIHNF KRKGGIGGYS AGERIVDIIA TDIQTKELQK QITKIQNFRV YYRDSRDPLW KGPAKLLWKG [960]
Bpol.LScot ..... [960]
Bpol.MMcot ..... [960]

[      970      980      990      1000 ]
[      .      .      .      .]
Bpol.mrca  EGAVVIQDNS DIKVVPRRKA KIIRDYGKQM AGDDCVASRQ DED [1003]
Bpol.LScot ..... [1003]
Bpol.MMcot ..... [1003]

```

Figure 22
Comparison of Clade B rev Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Brev.mrca	MAGRSGDSDE	ELLKTVRLIK	FLYQSNPPPS	PEGTRQARRN	RRRWRRERQR	QIRSISERIL	STYLGRSAEP	VPLQLPPLER	[80]
Brev.LScotW..P...	[80]
Brev.MMcotW..P...	[80]

[90	100	110]	
[
Brev.mrca	LTLDCSEDCG	TSGTQGVGSP	QILVESPAVL	ESGTKE	[116]
Brev.LScotN....T..	[116]
Brev.MMcot	[116]

Figure 23
Comparison of Clade B tat Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Btat.mrca	MEPVDPRLEP	WKHPGSQPKT	ACTNCYCKKC	CYHCQVCFIT	KGLGISYGRK	KRRQRRRPPQ	GSQTHQVSL	KQPASQPRGD	{80]
Btat.LScotF.....A..	D.....	{80]
Btat.MMcotF.....A..	D.....	{80]

[90	100]	
[
Btat.mrca	PTGPKESKKK	VERETETDPV	D {101]
Btat.LScot {101]
Btat.MMcot {101]

Figure 24
Comparison of Clade B vif Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Bvif.mrca	MENRWQVMIV	WQVDRMRIRT	WKSILVKHHMY	ISKKAKGWFY	RHHYESTHPR	ISSEVHIPLG	DARLVIKTYW	GLHTGEREWH	[80]
Bvif.LScotR.....T...D..	[80]
Bvif.MMcotR.....T...D..	[80]
[90	100	110	120	130	140	150	160]	
[
Bvif.mrca	LGQGVSIWR	KRRYSTQVDP	GLADQLIHLY	YFDCFSESAI	RNAILGHIVS	PRCEYQAGHN	KVGSLOQYLAL	TALITPKKIK	[160]
Bvif.LScotK.....	D.....	A.....	[160]
Bvif.MMcotK.....	D.....	A.....	[160]
[170	180	190]						
[
Bvif.mrca	PPLPSVRKLT	EDRWNKPKQT	KGHRGSHTMN	GH					[192]
Bvif.LScotT...					[192]
Bvif.MMcotT...					[192]

Figure 25
Comparison of Clade B vpr Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Bvpr.mrca	MEQAPEDQGP	QREPYNEWTL	ELLEELKSEA	VRHFPRLWLH	SLGQHIYETY	GDTWAGVEAI	IRILQQLLFI	HFRIGCQHSR	[80]
Bvpr.LScotI.....R...	[80]
Bvpr.MMcotI.....	[80]
[90]							
[
Bvpr.mrca	IGITRQRRAR	NGASRS							[96]
Bvpr.LScot							[96]
Bvpr.MMcot							[96]

Figure 26

```

[          10          20          30          40          50          60          70          80]
[          .          .          .          .          .          .          .          .]
Bvpu.mrca      MQPLEILAIV ALVVAAILAI VVWTIVFIEY RKILRQKID RLIDRIRERA EDSGNESEGD QEELSALVEM GHHPWDVDD [80]
Bvpu.LScot     ...Q.....I.. .. ... .. ... .. ... .. ... .. ... .. ... .. ... .. ... .. ... .. [80]
Bvpu.MMcot     ...Q.....I.. .. ... .. ... .. ... .. ... .. ... .. ... .. ... .. ... .. ... .. [80]

[          ]
[          ]
Bvpu.mrca      L      [81]
Bvpu.LScot     .      [81]
Bvpu.MMcot     .      [81]

```


[890	900	910	920	930	940	950	960]	
[.]	
Cgag.mrca	ATGTAGACCG	GTTCTTTAAA	ACTTTAAGAG	CTGAACAAGC	TACACAAGAT	GTAAAAAATT	GGATGACAGA	CACCTTGTTG	[957]
Cgag.LScot	[957]
Cgag.MMcot	[960]
[970	980	990	1000	1010	1020	1030	1040]	
[.]	
Cgag.mrca	GTCCAAAATG	CGAACCCAGA	TTGTAAGACC	ATTTTAAGAG	CATTAGGACC	AGGGGCTACA	CTAGAAGAAA	TGATGACAGC	[1037]
Cgag.LScot	T.....	[1037]
Cgag.MMcot	T.....	[1040]
[1050	1060	1070	1080	1090	1100	1110	1120]	
[.]	
Cgag.mrca	ATGTCAGGGA	GTGGGAGGAC	CTAGCCATAA	AGCAAGAGTT	TTGGCTGAGG	CAATGAGCCA	AGCAAACAAT	ACAAACATAA	[1117]
Cgag.LScotG...C..G	[1117]
Cgag.MMcotG...C..G	[1120]
[1130	1140	1150	1160	1170	1180	1190	1200]	
[.]	
Cgag.mrca	TGATGCAGAG	AGGCAATTTT	AAGGGCCCTA	GAAGAATTGT	TAAATGTTTC	AACTGTGGCA	AGGAAGGACA	CATAGCCAGA	[1197]
Cgag.LScotA.....	..A.....	A.....G..	[1197]
Cgag.MMcotA.....	..A.....	A.....G..	[1200]
[1210	1220	1230	1240	1250	1260	1270	1280]	
[.]	
Cgag.mrca	AATTGCAGGG	CCCCTAGGAA	AAAGGGCTGT	TGGAAATGTG	GAAAGGAAGG	ACACCAAATG	AAAGACTGTA	CTGAGAGGCA	[1277]
Cgag.LScotA.....	[1277]
Cgag.MMcotA.....	[1280]
[1290	1300	1310	1320	1330	1340	1350	1360]	
[.]	
Cgag.mrca	GGCTAATTTT	TTAGGGAAAA	TTTGGCCTTC	CCACAAGGGG	AGGCCAGGGA	ATTTCTTCA	GAGCAGACCA	GAGCCAACAG	[1357]
Cgag.LScot	[1357]
Cgag.MMcot	[1360]
[1370	1380	1390	1400	1410	1420	1430	1440]	
[.]	
Cgag.mrca	CCCCACCAGC	AGAGAGCTTC	AGGTTTCGAGG	AGACAACCCC	CGCTCCGAAG	CAGGAGCCGA	AAGACAGGGA	ACCCTTAACT	[1437]
Cgag.LScot	[1437]
Cgag.MMcot	[1440]
[1450	1460	1470	1480]					
[.	.	.	.]					
Cgag.mrca	TCCCTCAAAT	CACTCTTTGG	CAGCGACCCC	TTGTCTCAAT	AA				[1479]
Cgag.LScot				[1479]
Cgag.MMcot				[1482]

Figure 28
Comparison of Clade C env Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Cgp160.mrca	ATGAGAGTGA	TGGGGATACA	GAGGAATTGT	CAACAATGGT	GGATATGGGG	CATCTTAGGC	TTTTGGATGT	TAATGATTTG	[80]
Cgp160.LScotG.....T.....	[80]
Cgp160.MMcotG.....T.....	[80]
[90	100	110	120	130	140	150	160]	
[
Cgp160.mrca	TAGTGTGGTG	GGGAACTTGT	GGGTCACAGT	CTATTATGGG	GTACCTGTGT	GGAAAGAAGC	AAAACTACT	CTATTTTGTG	[160]
Cgp160.LScot	..A.....C....	[160]
Cgp160.MMcot	..A.....C....	[160]
[170	180	190	200	210	220	230	240]	
[
Cgp160.mrca	CATCAGATGC	TAAAGCATAT	GAGAGAGAAG	TGCATAATGT	CTGGGCTACA	CATGCCTGTG	TACCCACAGA	CCCCAACCCA	[240]
Cgp160.LScotA.....	[240]
Cgp160.MMcotA.....	[240]
[250	260	270	280	290	300	310	320]	
[
Cgp160.mrca	CAAGAAATGG	TTTTGGAAAA	TGTAACAGAA	AATTTTAACA	TGTGGAAAAA	TGACATGGTG	GATCAGATGC	ATGAGGATAT	[320]
Cgp160.LScot	[320]
Cgp160.MMcot	[320]
[330	340	350	360	370	380	390	400]	
[
Cgp160.mrca	AATCAGTTTA	TGGGATCAAA	GCCTAAAGCC	ATGTGTAAAG	TTGACCCAC	TCTGTGTCAC	TTTAACTGT	ACTAATGTTA	[400]
Cgp160.LScotT...G.....	[400]
Cgp160.MMcotT...G.....	[400]
[410	420	430	440	450	460	470	480]	
[
Cgp160.mrca	ATAATACTAA	TAATACCAAT	AGTACCATGA	ATGGAGAAAT	GAAAAATTGC	TCTTTCAATA	TAACCACAGA	AATAAGAGAT	[480]
Cgp160.LScot	..GC...C..	..C.....	..A.....	..G.....	..A.....G C...	[480]
Cgp160.MMcot	..C...C..	..C.....	..A.....	..A.....	..A.....GC.....	[480]
[490	500	510	520	530	540	550	560]	
[
Cgp160.mrca	AAGAAGAAGA	AAGAATATGC	ACTTTTTTAT	AGACTTGATA	TAGTACCACT	TAATGAAAAT	AATAACAATA	CTAGTGAATA	[560]
Cgp160.LScotAC...TG....G...G....TG..	[560]
Cgp160.MMcotA....G....G...TG..	[560]
[570	580	590	600	610	620	630	640]	
[
Cgp160.mrca	TAGATTAATA	AATTGTAATA	CCTCAGCCAT	AACACAAGCC	TGTCCAAAGG	TCTCTTTTGA	CCCAATTTCCT	ATACATTATT	[640]
Cgp160.LScot	[640]
Cgp160.MMcot	[640]
[650	660	670	680	690	700	710	720]	
[
Cgp160.mrca	GTGCTCCAGC	TGGTTATGCG	ATTCTAAAGT	GTAATAATAA	GACATTCAAT	GGAACAGGAC	CATGCAAAAA	TGTCAGCACA	[720]
Cgp160.LScotT..	[720]
Cgp160.MMcotT..	[720]
[730	740	750	760	770	780	790	800]	
[
Cgp160.mrca	GTACAATGTA	CACATGGAAT	TAAGCCAGTG	GTATCAACTC	AACTACTGTT	AAATGGTAGT	CTAGCAGAAG	AAGAGATAAT	[800]
Cgp160.LScotC	[800]
Cgp160.MMcotC	[800]
[810	820	830	840	850	860	870	880]	
[
Cgp160.mrca	AATTAGATCT	GAAAATCTGA	CAAACAATGC	CAAAACAATA	ATAGTACAGC	TTAATGAATC	TGTAGAAATT	GTGTGTACAA	[880]
Cgp160.LScotTT.	[880]
Cgp160.MMcotT.	[880]

[890	900	910	920	930	940	950	960]	
[]
Cgpl60.mrca	GACCCAACAA	TAATACAAGA	AAAAGTATGA	GGATAGGACC	AGGACAAACA	TTCTATGCAA	CAGGAGACAT	AATAGGAGAT	[960]
Cgpl60.LScotA.....C.....	[960]
Cgpl60.MMcotA.....C.....	[960]
[970	980	990	1000	1010	1020	1030	1040]	
[]
Cgpl60.mrca	ATAAGACAAG	CACATTGTAA	CATTAGTGGG	AGGGAATGGA	ATAACACTTT	ACAACAGGTA	GCTGAAAAAT	TAAGAAAACA	[1040]
Cgpl60.LScotA.....	GA.....A.....AG.....G.A.....GA.G.....	[1040]
Cgpl60.MMcotA.....	GA.....A.....AG.....G.A.....GA.G.....	[1040]
[1050	1060	1070	1080	1090	1100	1110	1120]	
[]
Cgpl60.mrca	CTTCCCTAAT	AAAACAATAA	AATTTGCACC	ATCCTCAGGA	GGGGACCTAG	AAATTACAAC	ACATAGCTTT	AATTGTAGAG	[1120]
Cgpl60.LScotA.....	[1120]
Cgpl60.MMcotA.....	[1120]
[1130	1140	1150	1160	1170	1180	1190	1200]	
[]
Cgpl60.mrca	GAGAATTTTT	CTATTGCAAT	ACATCAAAAC	TGTTTAATAG	TACATACAAT	AGTACAAATA	GTACAAATTC	AACCATCACA	[1200]
Cgpl60.LScot	G.....A.....	[1200]
Cgpl60.MMcot	G.....	[1200]
[1210	1220	1230	1240	1250	1260	1270	1280]	
[]
Cgpl60.mrca	CTCCCATGCA	GAATAAAACA	AATTATAAAC	ATGTGGCAGG	GGGTAGGACA	AGCAATGTAT	GCCCCCTCCA	TTGCAGGAAA	[1280]
Cgpl60.LScot	A.....	G.....	[1280]
Cgpl60.MMcotA.....	A.....	G.....	[1280]
[1290	1300	1310	1320	1330	1340	1350	1360]	
[]
Cgpl60.mrca	CATAACATGT	AAATCAAATA	TCACAGGACT	ACTATTGACA	CGTGATGGAG	GAAAAAATGA	AACTAATGAA	ACTGAGACAT	[1360]
Cgpl60.LScotGT.....A.....	C..A...A.C	..A...T..	[1360]
Cgpl60.MMcotGT.....	C..A...C	..A...T..	[1360]
[1370	1380	1390	1400	1410	1420	1430	1440]	
[]
Cgpl60.mrca	TCAGACCTGG	AGGAGGAGAT	ATGAGGGACA	ATTGGAGAAG	TGAATTATAT	AAATATAAAG	TAGTAGAAAT	TAAACCATTA	[1440]
Cgpl60.LScot	G.....	...G....G	[1440]
Cgpl60.MMcot	G.....	...G....G	[1440]
[1450	1460	1470	1480	1490	1500	1510	1520]	
[]
Cgpl60.mrca	GGAGTAGCAC	CCACTAAGGC	AAAAAGGAGA	GTGGTGGAGA	GAGAAAAAAG	AGCAGTGGGA	CTAGGAGCTG	TGTTCTTGG	[1520]
Cgpl60.LScot	...A.....	A.....	[1520]
Cgpl60.MMcot	...A.....	A.....	[1520]
[1530	1540	1550	1560	1570	1580	1590	1600]	
[]
Cgpl60.mrca	GTTCTTGGGA	GCAGCAGGAA	GCACTATGGG	CGCAGCGTCA	ATAACGCTGA	CGGTACAGGC	CAGACAATTA	TTGTCTGGTA	[1600]
Cgpl60.LScotG.....G.....	[1600]
Cgpl60.MMcotG.....	[1600]
[1610	1620	1630	1640	1650	1660	1670	1680]	
[]
Cgpl60.mrca	TAGTGCAACA	GCAAAGCAAT	TTGCTGAGGG	CTATAGAGGC	GCAACAGCAT	ATGTTGCAAC	TCACAGTCTG	GGGCATTAAG	[1680]
Cgpl60.LScotG.....	[1680]
Cgpl60.MMcotG.....	[1680]
[1690	1700	1710	1720	1730	1740	1750	1760]	
[]
Cgpl60.mrca	CAGCTCCAGG	CAAGAGTCCT	GGCTATGGAA	AGATACCTAA	AGGATCAACA	GCTCCTAGGG	ATTTGGGGCT	GCTCTGGAAA	[1760]
Cgpl60.LScotA.....A.....	[1760]
Cgpl60.MMcotA.....A.....	[1760]

[1770	1780	1790	1800	1810	1820	1830	1840]	
[
Cgpl60.mrca	ACTCATCTGC	ACCACTGCTG	TGCCTTGGA	CTCTAGTTGG	AGTAATAAAT	CTCAAGATGA	TATTTGGGAT	AACATGACCT	[1840]
Cgpl60.LScotA..	[1840]
Cgpl60.MMcotG..	[1840]
[1850	1860	1870	1880	1890	1900	1910	1920]	
[
Cgpl60.mrca	GGATGGAGTG	GGATAGAGAA	ATTAACAATT	ACACAGACAC	AATATACAGG	TTGCTTGAAG	AATCGCAAAA	CCAGCAGGAA	[1920]
Cgpl60.LScotC....GT....C.....	[1920]
Cgpl60.MMcotC....GT....C.....	[1920]
[1930	1940	1950	1960	1970	1980	1990	2000]	
[
Cgpl60.mrca	AAAAATGAAC	AAGATTTATT	GGCATTGGAC	AGTTGGGAAA	ATCTGTGGAA	TTGGTTTGAC	ATATCAAATT	GGCTGTGGTA	[2000]
Cgpl60.LScot	C.....AC.AA...A.....	[2000]
Cgpl60.MMcot	C.....AC.AA...A.....	[2000]
[2010	2020	2030	2040	2050	2060	2070	2080]	
[
Cgpl60.mrca	TATAAAATA	TTCATAATGA	TAGTAGGAGG	CTTGATAGGT	TTAAGAATAA	TTTTTGCTGT	GCTTTCTATA	GTAAATAGAG	[2080]
Cgpl60.LScotG.....	[2080]
Cgpl60.MMcotG.....	[2080]
[2090	2100	2110	2120	2130	2140	2150	2160]	
[
Cgpl60.mrca	TTAGGCAGGG	ATACTCACCT	TTGTCGTTTC	AGACCCTTAC	CCCAAACCCG	AGGGGACCCG	ACAGGCTCGA	AAGAATCGAA	[2160]
Cgpl60.LScotG.....	[2160]
Cgpl60.MMcotG.....	[2160]
[2170	2180	2190	2200	2210	2220	2230	2240]	
[
Cgpl60.mrca	GAAGAAGGTG	GAGAGCAAGA	CAGAGACAGA	TCCATTGCGAT	TAGTGAGCGG	ATTCTTAGCA	CTTGCCTGGG	ACGACCTGCG	[2240]
Cgpl60.LScot	[2240]
Cgpl60.MMcot	[2240]
[2250	2260	2270	2280	2290	2300	2310	2320]	
[
Cgpl60.mrca	GAGCCTGTGC	CTCTTCAGCT	ACCACCGCTT	GAGAGACTTC	ATCTTGATTG	CAGCGAGGAC	TGTGGAACCT	CTGGGACGCA	[2320]
Cgpl60.LScotA..A...G.G.AG.	G.....	[2320]
Cgpl60.MMcotA..A...G.G.AG.	G.....	[2320]
[2330	2340	2350	2360	2370	2380	2390	2400]	
[
Cgpl60.mrca	GCAGTCTCAG	GGGACTACAG	AGGGGGTGGG	AAGCCCTTAA	ATATCTGGGA	AGTCTTGTGC	AGTATTGGGG	TCAGGAGCTA	[2400]
Cgpl60.LScot	G.....T.....	[2400]
Cgpl60.MMcot	G.....T.....	[2400]
[2410	2420	2430	2440	2450	2460	2470	2480]	
[
Cgpl60.mrca	AAAAAGAGTG	CTATTAGTCT	GCTTGATACC	ATAGCAATAG	CAGTAGCTGA	AGGGACAGAT	AGGATTATAG	AAGTAGTACA	[2480]
Cgpl60.LScotA.....T..A....	[2480]
Cgpl60.MMcotA.....T..A....	[2480]
[2490	2500	2510	2520	2530	2540	2550]		
[
Cgpl60.mrca	AAGAGCTTGT	AGAGCTATCC	TCAACATACC	TAGAAGAATA	AGACAGGGCT	TTGAAGCAGC	TTTGCAATAA		[2550]
Cgpl60.LScotAT....	G.....		[2550]
Cgpl60.MMcotAT....	G.....		[2550]

Figure 30
Comparison of Clade C pol Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.
Cpol.mrca	TTTTTTAGGG	AAAATTGGC	CTTCCCACAA	GGGGAGGCCA	GGAATTTC	TTCAGAGCAG	ACCAGAGCCA	ACAGCCCCAC	[80]
Cpol.LScot	[80]
Cpol.MMcot	[80]
[90	100	110	120	130	140	150	160]	
[.
Cpol.mrca	CAGCAGAGAG	CTTCAGGTTT	GAGGAGACAA	CCCCCGCTCC	GAAGCAGGAG	CCGAAAGACA	GGAACCCTT	AACTTCCTC	[160]
Cpol.LScot	[160]
Cpol.MMcotT.....	[160]
[170	180	190	200	210	220	230	240]	
[.
Cpol.mrca	AAATCACTCT	TTGGCAGCGA	CCCCTGTCT	CAATAAAAGT	AGGGGGCCAG	ATAAAGGAAG	CTCTATTAGA	TACAGGAGCA	[240]
Cpol.LScotG.....C.....	C.....	[240]
Cpol.MMcotA.....A.....	C.....G.....C.....	C.....	[240]
[250	260	270	280	290	300	310	320]	
[.
Cpol.mrca	GATGATACAG	TATTAGAAGA	CATAAATTG	CCAGGAAAAT	GGAAACCAAA	AATGATAGGG	GGAATTGGAG	GTTTTATCAA	[320]
Cpol.LScot	A.....A.....	[320]
Cpol.MMcot	A.....A.....	[320]
[330	340	350	360	370	380	390	400]	
[.
Cpol.mrca	AGTAAGACAG	TATGATCAAA	TACTTATAGA	AATTTGTGGA	AAAAAGGCTA	TAGGTACAGT	ATTAGTAGGA	CCTACACCTG	[400]
Cpol.LScot	[400]
Cpol.MMcotC.....	[400]
[410	420	430	440	450	460	470	480]	
[.
Cpol.mrca	TCAACATAAT	TGGAAGAAAT	ATGTTGACTC	AGCTTGCTTG	CACTCTAAAT	TTTCCAATTA	GTCCTATTGA	AACTGTACCA	[480]
Cpol.LScotA.....A.....C.....	[480]
Cpol.MMcotA.....C.....	[480]
[490	500	510	520	530	540	550	560]	
[.
Cpol.mrca	GTAAATTAAT	AGCCAGGAAT	GGATGGCCCA	AAGGTTAAAC	AATGGCCATT	GACAGAAGAG	AAAATAAAAG	CATTAACAGC	[560]
Cpol.LScot	[560]
Cpol.MMcotC.....	[560]
[570	580	590	600	610	620	630	640]	
[.
Cpol.mrca	AATTTGTGAA	GAAATGGAAA	AGGAAGGAAA	AATTACAAAA	ATTGGGCCTG	AAAATCCATA	TAACACTCCA	GTATTTGCCA	[640]
Cpol.LScotG.....	[640]
Cpol.MMcotG.....	[640]
[650	660	670	680	690	700	710	720]	
[.
Cpol.mrca	TAAAAAAGAA	GGACAGTACT	AAGTGGAGAA	AATTAGTAGA	TTTCAGAGAA	CTCAATAAAA	GAACTCAAGA	CTTCTGGGAA	[720]
Cpol.LScotG.....T.....	[720]
Cpol.MMcotG.....T.....	[720]
[730	740	750	760	770	780	790	800]	
[.
Cpol.mrca	GTTCAATTAG	GAATACCACA	CCCAGCAGGG	TTAAAAAAGA	AAAAATCAGT	AACAGTACTG	GATGTGGGGG	ATGCATATTT	[800]
Cpol.LScot	G.....	[800]
Cpol.MMcot	G.....	G.....	[800]
[810	820	830	840	850	860	870	880]	
[.
Cpol.mrca	TTCAGTTCCT	TTAGATGAAG	ACTTCAGGAA	ATATACTGCA	TTCACCATAC	CTAGTATAAA	CAATGAAACA	CCAGGGATTA	[880]
Cpol.LScot	G.....	[880]
Cpol.MMcot	G.....	[880]

[890	900	910	920	930	940	950	960]	
[.]								
Cpol.mrca	GATATCAATA	TAATGTGCTT	CCACAGGGAT	GGAAAGGATC	ACCAGCAATA	TTCCAGAGTA	GCATGACAAA	AATCTTAGAG	[960]
Cpol.LScot	[960]
Cpol.MMcot	[960]
[970	980	990	1000	1010	1020	1030	1040]	
[.]								
Cpol.mrca	CCCTTTAGGG	CACAAAACCC	AGAAATAGTT	ATCTATCAAT	ACATGGATGA	CTTGTATGTA	GGATCTGACT	TAGAAATAGG	[1040]
Cpol.LScotT..CT.....	[1040]
Cpol.MMcotT..T.....	[1040]
[1050	1060	1070	1080	1090	1100	1110	1120]	
[.]								
Cpol.mrca	GCAACATAGA	GCAAAAATAG	AGGAGTTAAG	AGAACATCTA	TTGAAATGGG	GATTTACCAC	ACCAGACAAG	AAACATCAGA	[1120]
Cpol.LScotA..G...	[1120]
Cpol.MMcotG.....	[1120]
[1130	1140	1150	1160	1170	1180	1190	1200]	
[.]								
Cpol.mrca	AAGAACCCCC	ATTTCTTTGG	ATGGGGTATG	AACTCCATCC	TGACAAATGG	ACAGTACAGC	CTATACAGCT	GCCAGAAAAG	[1200]
Cpol.LScot	[1200]
Cpol.MMcot	[1200]
[1210	1220	1230	1240	1250	1260	1270	1280]	
[.]								
Cpol.mrca	GATAGCTGGA	CTGTCAATGA	TATACAGAAG	TTAGTGGGAA	AATTAACTG	GGCAAGTCAG	ATTTACCCAG	GGATTAAAGT	[1280]
Cpol.LScot	[1280]
Cpol.MMcot	[1280]
[1290	1300	1310	1320	1330	1340	1350	1360]	
[.]								
Cpol.mrca	AAGGCAACTG	TGTAAACTCC	TTAGGGGAGC	CAAAGCACTA	ACAGACATAG	TACCACTGAC	TGAAGAAGCA	GAATTAGAAT	[1360]
Cpol.LScotTG..A..	[1360]
Cpol.MMcotA..	[1360]
[1370	1380	1390	1400	1410	1420	1430	1440]	
[.]								
Cpol.mrca	TGGCAGAGAA	CAGGGAAATT	CTAAAAGAAC	CAGTACATGG	AGTATATTAT	GACCCATCAA	AAGACTTAAT	AGCTGAAATA	[1440]
Cpol.LScotG..	[1440]
Cpol.MMcot	[1440]
[1450	1460	1470	1480	1490	1500	1510	1520]	
[.]								
Cpol.mrca	CAGAAACAGG	GGCATGACCA	ATGGACATAT	CAAATTTACC	AAGAACCATT	CAAAAATCTG	AAAACAGGAA	AGTATGCAAA	[1520]
Cpol.LScotG.	[1520]
Cpol.MMcotG.	[1520]
[1530	1540	1550	1560	1570	1580	1590	1600]	
[.]								
Cpol.mrca	AATGAGGTCT	GCCCACACTA	ATGATGTAAA	ACAATTAACA	GAAGCAGTGC	AAAAAATAGC	CATGGAAAGC	ATAGTAATAT	[1600]
Cpol.LScotA..G.....G.....	[1600]
Cpol.MMcotA..G.....G.....	[1600]
[1610	1620	1630	1640	1650	1660	1670	1680]	
[.]								
Cpol.mrca	GGGGAAAGAC	TCCTAAATTT	AGACTACCCA	TCCAAAAAGA	AACATGGGAG	ACATGGTGGA	CAGACTATTG	GCAAGCCACC	[1680]
Cpol.LScot	[1680]
Cpol.MMcotT.....G.....G.....	[1680]
[1690	1700	1710	1720	1730	1740	1750	1760]	
[.]								
Cpol.mrca	TGGATTCCTG	AGTGGGAGTT	TGTTAATACC	CCTCCCCTAG	TAAAATTATG	GTACCAGCTA	GAAAAAGAAC	CCATAGCAGG	[1760]
Cpol.LScotGG	[1760]
Cpol.MMcotGG	[1760]

[[1770	1780	1790	1800	1810	1820	1830	1840]	
[[.]
Cpol.mrca		AGCAGAAACT	TTCTATGTAG	ATGGGGCAGC	TAATAGGGAA	ACTAAACTAG	GAAAAGCAGG	GTATGTTACT	GACAAAGGAA	[1840]
Cpol.LScot	A.....A.....G.....	[1840]
Cpol.MMcot	A.....A.....G.....	[1840]
[[1850	1860	1870	1880	1890	1900	1910	1920]	
[[.]
Cpol.mrca		GACAGAAAGT	TGTTTCTCTA	ACTGAAACAA	CAAATCAGAA	GACTGAATTA	CAAGCAATTC	AGCTAGCTTT	GCAGGATTCA	[1920]
Cpol.LScot		.G.....A.A.....	[1920]
Cpol.MMcot		.G.....A.	[1920]
[[1930	1940	1950	1960	1970	1980	1990	2000]	
[[.]
Cpol.mrca		GGATCAGAAG	TAAACATAGT	AACAGACTCA	CAATATGCAT	TAGGAATCAT	TCAAGCACAA	CCAGATAAGA	GTGAATCAGA	[2000]
Cpol.LScot	G.....	[2000]
Cpol.MMcot	G.....	[2000]
[[2010	2020	2030	2040	2050	2060	2070	2080]	
[[.]
Cpol.mrca		GTTAGTCAAT	CAAATAATAG	AGCAGTTAAT	AAAAAAGGAA	AAGGTCTACC	TGTCATGGGT	ACCAGCACAT	AAAGGAATTG	[2080]
Cpol.LScot	C.....A..A.....G.....	[2080]
Cpol.MMcot	A.....G.....	[2080]
[[2090	2100	2110	2120	2130	2140	2150	2160]	
[[.]
Cpol.mrca		GAGGAAATGA	ACAAGTAGAT	AAATTAGTAA	GTTCTGGAAT	CAGGAAAAGTG	CTGTTTCTAG	ATGGAATAGA	TAAAGCTCAA	[2160]
Cpol.LScot	AG.....G.....	[2160]
Cpol.MMcot	AG.....G.....	[2160]
[[2170	2180	2190	2200	2210	2220	2230	2240]	
[[.]
Cpol.mrca		GAAGAACATG	AAAAATATCA	CAGCAATTGG	AGAGCAATGG	CTAGTGAGTT	TAATCTGCCA	CCCATAGTAG	CAAAAGAAAT	[2240]
Cpol.LScot	G.....G.....	[2240]
Cpol.MMcot	G.....	[2240]
[[2250	2260	2270	2280	2290	2300	2310	2320]	
[[.]
Cpol.mrca		AGTAGCTAGC	TGTGATAAAT	GTCAGCTAAA	AGGGGAAGCC	ATGCATGGAC	AAGTAGACTG	TAGTCCAGGG	ATATGGCAAT	[2320]
Cpol.LScot	A.....	[2320]
Cpol.MMcot	A.....	[2320]
[[2330	2340	2350	2360	2370	2380	2390	2400]	
[[.]
Cpol.mrca		TAGATTGTAC	ACATTTAGAA	GGAAAAGTTA	TCCTGGTAGC	AGTCCATGTA	GCCAGTGGCT	ACATAGAAGC	AGAAGTTATC	[2400]
Cpol.LScot	A.C.....G.....	[2400]
Cpol.MMcot	A.C.....G.....	[2400]
[[2410	2420	2430	2440	2450	2460	2470	2480]	
[[.]
Cpol.mrca		CCAGCAGAAA	CAGGACAGGA	AACAGCATAC	TTTATATTAA	AATTAGCAGG	AAGATGGCCA	GTAAAAGTAA	TACATACAGA	[2480]
Cpol.LScot	A..C.....C.....	[2480]
Cpol.MMcot	A..C.....C.....	[2480]
[[2490	2500	2510	2520	2530	2540	2550	2560]	
[[.]
Cpol.mrca		CAATGGCAGC	AATTTACCA	GTGCTGCAGT	TAAGGCAGCC	TGTTGGTGGG	CAGGTATCCA	ACAGGAATTT	GGAATTCCT	[2560]
Cpol.LScot	T.....	[2560]
Cpol.MMcot	T.....A.....					

```

[          2730          2740          2750          2760          2770          2780          2790          2800]
[          .          .          .          .          .          .          .          .]
Cpol.mrca  AGGGGAAAGA ATAATAGACA TAATAGCAAC AGACATACAA ACTAAAGAAT TACAAAAACA AATTATAAAA ATTCAAAATT [2800]
Cpol.LScot  ..... [2800]
Cpol.MMcot  ..... [2800]

[          2810          2820          2830          2840          2850          2860          2870          2880]
[          .          .          .          .          .          .          .          .]
Cpol.mrca  TTCGGGTTTA TTACAGAGAC AGCAGAGACC CTGTTTGGAA AGGACCAGCC AAACACTCTT GGAAAGGTGA AGGGGCAGTA [2880]
Cpol.LScot  ..... A..... [2880]
Cpol.MMcot  ..... A..... [2880]

[          2890          2900          2910          2920          2930          2940          2950          2960]
[          .          .          .          .          .          .          .          .]
Cpol.mrca  GTAATACAAG ACAATAGTGA CATAAAGGTA GTACCAAGGA GGAAAGCAAA GATCATTAGG GATTATGGAA AACAGATGGC [2960]
Cpol.LScot  ..... T..... A..... A. ..C..... [2960]
Cpol.MMcot  ..... T..C..... A..... A. ..C..... [2960]

[          2970          2980          2990          3000]
[          .          .          .          .]
Cpol.mrca  AGGTGCTGAT TGTGTGGCAG GTAGACAGGA TGAAGATTAG [3000]
Cpol.LScot  ..... [3000]
Cpol.MMcot  ..... [3000]

```

Figure 31
Comparison of Clade C rev Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.]
Crev.mcra	ATGGCAGGAA	GAAGCGGAGA	CAGCGACGAA	GCGCTCCTCC	AAGCAGTGAG	GATCATCAAA	ATCCTATATC	AAAGCAACCC	[80]
Crev.LScotT.....	[80]
Crev.MMcot	[80]
[90	100	110	120	130	140	150	160]	
[.]
Crev.mcra	TTACCCCAAA	CCCAGAGGGA	CCCGACAGGC	TCGAAGGAAT	CGAAGAAGAA	GGTGGAGAGC	AAGACAGAGA	CAGATCCATT	[160]
Crev.LScotG.A....	[160]
Crev.MMcotG.A....	[160]
[170	180	190	200	210	220	230	240]	
[.]
Crev.mcra	CGATTAGTGA	GCGGATTCTT	AGCACTTGCC	TGGGACGACC	TGCGGAGCCT	GTGCCTCTTC	AGCTACCACC	GCTTGAGAGA	[240]
Crev.LScotA.....	[240]
Crev.MMcotT...A.....	[240]
[250	260	270	280	290	300	310	320]	
[.]
Crev.mcra	CTTCATCTTG	ATTGCAGCGA	GGA CTGTGGA	ACTTCTGGGA	CGCAGCAGTC	TCAGGGGACT	ACAGAGGGGG	TGGGAAGCCC	[320]
Crev.LScotA...	G.GA.....	.AG.A....	[320]
Crev.MMcotA...	G.GA.....	.AG.A....	[320]
[330	340	350	360	370	380]			
[.]
Crev.mcra	TTAAATATCT	GGAAGCCTT	GTGCAGTATT	GGGGTCAGGA	GCTAAAAAAG	AGTGCTATTA	G		[381]
Crev.LScotG....T...T...	A.....			[381]
Crev.MMcotG....T...T...	A.....			[381]

Figure 32
Comparison of Clade C tat Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.]
Ctat.mrca	ATGGAGCCAG	TAGATCCTAA	CCTAGAGCCC	TGGAACCATC	CAGGAAGTCA	GCCTAAAACT	GCTTGTAATA	AATGTTATTG	[80]
Ctat.LScot	C.....	.G.....	[80]
Ctat.MMcot	C.....	.G.....	[80]
[90	100	110	120	130	140	150	160]	
[.]
Ctat.mrca	TAAAAAATGT	AGCTATCATT	GTCTAGTTTG	CTTTCTGACA	AAAGGCTTAG	GCATTTCCTA	TGGCAGGAAG	AAGCGGAGAC	[160]
Ctat.LScotC.C...A....	[160]
Ctat.MMcotC.C...A....	[160]
[170	180	190	200	210	220	230	240]	
[.]
Ctat.mrca	AGCGACGAAG	AGCTCCTCCA	AGCAGTGAGG	ATCATCAAAA	TCCTATATCA	AAGCAACCCT	TATCCCAAAC	CCGAGGGGAC	[240]
Ctat.LScot	C.....C.....	[240]
Ctat.MMcot	C.....C.....	[240]
[250	260	270	280	290	300]		
[.]		
Ctat.mrca	CCGACAGGCT	CGGAGGAATC	GAAGAAGAAG	GTGGAGAGCA	AGACAGAGAC	AGATCCGTGC	GATTAG		[306]
Ctat.LScotA....A.T.		[306]
Ctat.MMcotA.T.		[306]

Figure 33
Comparison of Clade C vif Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.]
Cvif.mrca	ATGGAAAACA	GATGGCAGGT	GCTGATTGTG	TGGCAGGTAG	ACAGGATGAA	GATTAGAACA	TGGAATAGTT	TAGTAAAACA	[80]
Cvif.LScotG..	[80]
Cvif.MMcotG..	[80]
[90	100	110	120	130	140	150	160]	
[.]
Cvif.mrca	CCATATGTAT	GTTTCAAGGA	GAGCTAAAGG	ATGGTTTTAT	AGACATCACT	ATGAAAGCAG	ACATCCAAAA	ATAAGTTCAG	[160]
Cvif.LScotT..CT.G.....	[160]
Cvif.MMcotT..CT.G.....	[160]
[170	180	190	200	210	220	230	240]	
[.]
Cvif.mrca	AAGTACACAT	CCCATTAGGG	GATGCTAGAT	TAGTAATAAA	AACATATTGG	GGTTTGCATA	CAGGAGAAAG	AGATTGGCAT	[240]
Cvif.LScotA.	[240]
Cvif.MMcot	[240]
[250	260	270	280	290	300	310	320]	
[.]
Cvif.mrca	TTGGGTCATG	GAGTCTCCAT	AGAATGGAGA	CTGAGAAGAT	ATAGCACACA	AGTAGACCCT	GGCCTGGCAG	ACCAACTAAT	[320]
Cvif.LScotT.G.....	[320]
Cvif.MMcotT.G.....	[320]
[330	340	350	360	370	380	390	400]	
[.]
Cvif.mrca	TCATATGCAT	TATTTTGATT	GTTTTCAGAG	CTCTGCCATA	AGGAAAGCCA	TATTAGGACA	TATAGTTAGC	CCTAGGTGTG	[400]
Cvif.LScotA.....C.....TT	[400]
Cvif.MMcotA.....C.....TT	[400]
[410	420	430	440	450	460	470	480]	
[.]
Cvif.mrca	ACTATCAAGC	AGGACATAAC	AAGGTAGGAT	CTCTACAATA	CTTGGCACTG	ACAGCATTAA	TAAAACCAAA	AAAGATAAAG	[480]
Cvif.LScotTG.	[480]
Cvif.MMcotTG.	[480]
[490	500	510	520	530	540	550	560]	
[.]
Cvif.mrca	CCACCTCTGC	CTAGTGTTAA	GAAATTAGTA	GAGGATAGAT	GGAACAAGCC	CCAGAAGACC	AGGGGCCACA	GAGGGAGCCA	[560]
Cvif.LScotGG..A...	[560]
Cvif.MMcotGG..A...	[560]
[570								
[.								
Cvif.mrca	TACAATGAAT	GGACACTAG							[579]
Cvif.LScot							[579]
Cvif.MMcot							[579]

Figure 34
Comparison of Clade C vpr Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.]
Cvpr.mrca	ATGGAACAAG	CCCCAGAAGA	CCAGGGGCCA	CAGAGGGAGC	CATACAATGA	ATGGACACTA	GAGCTTTTAG	AGGAACTTAA	[80]
Cvpr.LScotGAA.....C..	[80]
Cvpr.MMcotGAC..	[80]
[90	100	110	120	130	140	150	160]	
[.]
Cvpr.mrca	GCAGGAAGCT	GTCAGACATT	TTCCTAGACC	ATGGCTCCAT	AGCTTAGGAC	AACATATCTA	TGAAACCTAT	GGGGATACTT	[160]
Cvpr.LScotCT.....	[160]
Cvpr.MMcotC	[160]
[170	180	190	200	210	220	230	240]	
[.]
Cvpr.mrca	GGGCGGGAGT	TGAAGCTATA	ATAAGAATTC	TGCAACAAC	ACTGTTTATT	CATTTTCAGAA	TTGGGTGCCA	ACATAGCAGA	[240]
Cvpr.LScot	..A.A....	C.....C..A	G.....	[240]
Cvpr.MMcot	..A.....C..	G.....	[240]
[250	260	270	280	290]				
[.]
Cvpr.mrca	ATAGGCATTA	TTCGACAGAG	AAGAGCAAGA	AATGGAGCCA	GATAGTCCTA	A			[291]
Cvpr.LScotT	.G.....			[291]
Cvpr.MMcotG.....			[291]

Figure 35
Comparison of Clade C vpu Gene Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.]
Cvpu.mrca	ATGTTAGATT	TAATAGCAAG	AGTAGATTAT	AGATTAGGAG	TAGGAGCATT	GATAGTAGCA	CTAATCATAG	CAATAGTTGT	[80]
Cvpu.LScotC.....	[80]
Cvpu.MMcotC.....	[80]
[90	100	110	120	130	140	150	160]	
[.]
Cvpu.mrca	GTGGACCATA	GTATATATAG	AATATAGGAA	ATTGGTAAGA	CAAAGAAAAA	TAGACTGGTT	AATTAAAAGA	ATTAGGGAAA	[160]
Cvpu.LScotT.....	[160]
Cvpu.MMcotT.....	[160]
[170	180	190	200	210	220	230	240]	
[.]
Cvpu.mrca	GAGCAGAAGA	CAGTGGCAAT	GAGAGTGATG	GGGATACAGA	GGAATTGTCA	ACACTGGTGG	ATATGGGGCA	TCTTAGGCTT	[240]
Cvpu.LScotG.T..A.....	[240]
Cvpu.MMcotG.T..A..	...A.....	[240]
[250	260]							
[.	.]
Cvpu.mrca	TTGGATGTTA	ATGATTGTA A	[261]						
Cvpu.LScot	[261]						
Cvpu.MMcot	[261]						

Figure 36
Comparison of Clade C gag Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.]
Cgag.mrca	MGARASILRG	GKLDTWEEKIR	LRPGGKKHYM	IKHLVWASRE	LERFALNPGL	LETSEGCKQI	IKQLQPALQT	GTEELKSLYN	[80]
Cgag.LScot	L.....	M.....R....	[80]
Cgag.MMcot	L.....	M.....R....	[80]
[90	100	110	120	130	140	150	160]	
[.]
Cgag.mrca	TVATLYCVHQ	RIEVRDTKEA	LDKIEEEQNK	SQOKTQQAEE	-ADGKVSQNY	PIVQNQQGQM	VHQAI SPRTL	NAWVKVIEEK	[159]
Cgag.LScotE K	[159]
Cgag.MMcotE K	A.....	[160]
[170	180	190	200	210	220	230	240]	
[.]
Cgag.mrca	AFSPEVIPMF	TALSEGATPQ	DLNTMLNTVG	GHQAAMQMLK	DTINEEAAEW	DRLHPVHAGP	VAPGQMREPR	GSDIAGTTST	[239]
Cgag.LScot	[239]
Cgag.MMcot	[240]
[250	260	270	280	290	300	310	320]	
[.]
Cgag.mrca	LQEQIAWMTS	NPPIPVVDIY	KRWIILGLNK	IVRMYSVSI	LDIKQGPKEP	FRDYVDRFFK	TLRAEQATQD	VKNWMTDTLL	[319]
Cgag.LScotV.....	[319]
Cgag.MMcotV.....	[320]
[330	340	350	360	370	380	390	400]	
[.]
Cgag.mrca	VQNANPDCKT	ILRALGPGAT	LEEMMTACQG	VGGPSHKARV	LAEAMSQANN	TNIMMQRGNF	KGPRRIVKCF	NCGKEGHIAR	[399]
Cgag.LScotG.....S..K.....	[399]
Cgag.MMcotG.....S..K.....	[400]
[410	420	430	440	450	460	470	480]	
[.]
Cgag.mrca	NCRAPRKKGC	WKCCKEGHQM	KDCTERQANF	LGKIWPSHKG	RPGNFLQSRP	EPTAPPAESF	RFEETTPAPK	QEPKDREPLT	[479]
Cgag.LScot	[479]
Cgag.MMcot	[480]
[490]								
[.]
Cgag.mrca	SLKSLFGSDP	LSQ							[492]
Cgag.LScot							[492]
Cgag.MMcot							[493]

Figure 37
Comparison of Clade C gp160 Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.]
Cgp160.mrca	MRVMGIQRNC	QQWWIWGILG	FWMLMICSVV	GNLWVTVYYG	VPVWKEAKTT	LFCASDAKAY	EREVHNVWAT	HACVPTDPNP	[80]
Cgp160.LScot	...R..L...N..K.....	[80]
Cgp160.MMcot	...R..L...N..K.....	[80]
[90	100	110	120	130	140	150	160]	
[.]
Cgp160.mrca	QEMVLENVTE	NFNMWKNDMV	DQMHEDIISL	WDQSLKPCVK	LTPLCVTLNC	TNVNNTNNTN	STMNGEMKNC	SFNITTEIRD	[160]
Cgp160.LScot	S...A..T..	N..K..I...	...A.....	[160]
Cgp160.MMcot	S...T..T..	N..K..I...	...V...L..	[160]
[170	180	190	200	210	220	230	240]	
[.]
Cgp160.mrca	KKKKEYALFY	RLDIVPLNEN	NNNTSEYRLI	NCNTSAITQA	CPKVSFDPIP	IHYCAPAGYA	ILKCNNKTFN	GTGPCKNVST	[240]
Cgp160.LScot	..Q..V....S.S.N...	[240]
Cgp160.MMcotS.N...	[240]
[250	260	270	280	290	300	310	320]	
[.]
Cgp160.mrca	VQCTHGIPKV	VSTQLLLNGS	LAEEIIIRIS	ENLTNNAKTI	IVQLNESVEI	VCTRPNNNTR	KSMRIGPGQT	FYATGDIIGD	[320]
Cgp160.LScotV...	..H.....I.....	[320]
Cgp160.MMcotH.....I.....	[320]
[330	340	350	360	370	380	390	400]	
[.]
Cgp160.mrca	IRQAHCNISG	REWNNTLQQV	AEKLRKHFPN	KTIKFAPSSG	GDLEITTHSF	NCRGEFFYCN	TSKLPNSTYN	STNSTNSTIT	[400]
Cgp160.LScotE E...K...R.	GK..EE....E....G.....T...	[400]
Cgp160.MMcotE E...K...R.	GK..EE....E....G.....	[400]
[410	420	430	440	450	460	470	480]	
[.]
Cgp160.mrca	LPCRICKIIN	MWQGVGQAMY	APPIAGNITC	KSNITGLLLT	RDGGKNETNE	TETFRPGGGD	MRDNWRSELY	KYKVVVEIKPL	[480]
Cgp160.LScotE..R...VN..N	..I.....	[480]
Cgp160.MMcot	..Q.....E..R...VD..D	..I.....	[480]
[490	500	510	520	530	540	550	560]	
[.]
Cgp160.mrca	GVAPTAKARR	VVEREKRAVG	LGAVFLGFLG	AAGSTMGAAS	ITLTVQARQL	LSGIVQQQSN	LLRAIEAQQH	MLQLTVWGIK	[560]
Cgp160.LScot	..I.....I.....	[560]
Cgp160.MMcot	..I.....I.....	[560]
[570	580	590	600	610	620	630	640]	
[.]
Cgp160.mrca	QLQARVLAME	RYLKDQQLLG	IWGCSGKLIC	TTAVPWNSSW	SNKSQDDIWD	NMTWMEWDRE	INNYTDTIYR	LLEESQNQQE	[640]
Cgp160.LScot	...T....I.E....Q....	..S.....	...D.....	[640]
Cgp160.MMcot	...T....I.E....Q....	..S.....	...D.....	[640]
[650	660	670	680	690	700	710	720]	
[.]
Cgp160.mrca	KNEQDLLALD	SWENLWNWFD	ISNLWYIKI	FIMIVGGLIG	LRIIFAVLSI	VNRVRQGYSP	LSFQTLTPNP	RGPDRLERIE	[720]
Cgp160.LScot	Q..K.....	..K.....	..T.....G...	[720]
Cgp160.MMcot	Q..K.....	..K.....	..T.....G...	[720]
[730	740	750	760	770	780	790	800]	
[.]
Cgp160.mrca	EEGGEQDRDR	SIRLVSGFLA	LAWDDLRLSLC	LFSYHRLRDF	ILIAARTVEL	LGRSSLRGLQ	RGWEALKYLG	SLVQYWQEL	[800]
Cgp160.LScotV...A...L...	[800]
Cgp160.MMcotV...A...L...	[800]
[810	820	830	840					
[.]
Cgp160.mrca	KKSAISLLDT	IAIAVAEGTD	RIIEVVQRAC	RAILNIPPRI	RQGFEAALQ				[849]
Cgp160.LScotLI..I.	...R.....				[849]
Cgp160.MMcotLI..I.	...R.....				[849]

Figure 38
Comparison of Clade C nef Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.]
Cnrf.mrca	MGGKWSKSSI	VGWPAVRERI	RRTAPAAEGV	GAASQDLDKH	GALTSSNTAA	TNADCAWLEA	QEEE-EVGFP	VRPQVPLRPM		[79]
Cnrf.LScotE.....	N.....	.E.....		[80]
Cnrf.MMcotE.....	N.....	.E.....		[80]
[90	100	110	120	130	140	150	160]	
[.]
Cnrf.mrca	TYKGAVDLSF	FLKEKGGLEG	LIYSKKRQEI	LDLWVYHTQG	YFPDWQNYTP	GPGIRFPLTF	GWCFKLV PVD	PREVEEANEG		[159]
Cnrf.LScot	...F....V.Y...		[160]
Cnrf.MMcot	...F....V.Y...		[160]
[170	180	190	200]				
[.	.	.	.]				
Cnrf.mrca	ENNCLLHPMS	QHGMEDEDRE	VLKWKFDSLH	ARRHMARELH	PEYYKDC					[206]
Cnrf.LScot					[207]
Cnrf.MMcot					[207]


```

[      890      900      910      920      930      940      950      960]
[      .      .      .      .      .      .      .      .]
Cpol.mrca EHLKTAVQMA VFIHNFKRKG GIGGYSAGER IIDIIATDIQ TKELQKQIIK IQNFRVYYRD SRDPVWKGPA KLLWKGEHAV [960]
Cpol.LScot .....I..... [960]
Cpol.MMcot .....I..... [960]

[      970      980      990      ]
[      .      .      .      ]
Cpol.mrca VIQDNSDIKV VPRRKAKIIR DYGQMAGAD CVAGRQDED [999]
Cpol.LScot .....K..... [999]
Cpol.MMcot .....K..... [999]

```

Figure 40
Comparison of Clade C rev Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Crev.mcra	MAGRSGDSDE	ALLQAVRIIK	ILYQSNPYPK	PEGTRQARRN	RRRRWRARQR	QIHSISERIL	STCLGRPAEP	VPLQLPPLER	[80]
Crev.LScotK.I..	[80]
Crev1.MMcotK.F....I..	[80]
[90	100							
[
Crev.mcra	LHLDCSEDCG	TSGTQQSQGT	TEGVGSP						[107]
Crev.LScot	..IGD..SS.						[107]
Crev1.MMcot	..IGD..SS.						[107]

Figure 41
Comparison of Clade C tat Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[
Ctat.mrca	MEPVDPNLEP	WNHPGSQPKT	ACNKCYCKKC	SYHCLVCFLT	KGLGISYGRK	KRRQRRRAPP	SSEDHQNPIIS	KQPLSQTRGD	[80]
Ctat.LScot	P.....H.Q.S...P.....	[80]
Ctat.MMcot	P.....H.Q.S...P.....	[80]
[90	100]							
[
Ctat.mrca	PTGSEESKKK	VESKTETDPC	D						[101]
Ctat.LScotF	.						[101]
Ctat.MMcotF	.						[101]

Figure 42
Comparison of Clade C vif Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.]
Cvif.mrca	MENRWQVLIV	WQVDRMKIRT	WNSLVKHHMY	VSRRAKGWFY	RHHYESRHPK	ISSEVHIPLG	DARLVIKTYW	GLHTGERDWH	[80]
Cvif.LScotN...	V.....Q.....	[80]
Cvif.MMcotN...	V.....	[80]
[90	100	110	120	130	140	150	160]	
[.]
Cvif.mrca	LGHGVSIEWR	LRRYSTQVDP	GLADQLIHMH	YFDCFADSAI	RKAILGHIVS	PRCDYQAGHN	KVGSLOQLAL	TALIKPKKIK	[160]
Cvif.LScotI	[160]
Cvif.MMcotI	[160]
[170	180	190]						
[.	.	.]					
Cvif.mrca	PPLPSVKKL	EDRWNKPKT	RGHRGSHTMN	GH					[192]
Cvif.LScotR...R..N...	..					[192]
Cvif.MMcotR...R..N...	..					[192]

Figure 43
Comparison of Clade C vpr Protein Sequence Reconstructions

[10	20	30	40	50	60	70	80]	
[.]
Cvpr.mrca	MEQAPEDQGP	QREPYNEWTL	ELLEELKQEA	VRHFPRPWLH	SLGQHIYETY	GDTWAGVEAI	IRILQQLLFI	HFRIGCQHSR	[80]
Cvpr.LScotI.....Y....T...L	[80]
Cvpr.MMcotT...L	[80]
[90]							
[.]							
Cvpr.mrca	IGIIRQRRAR	NGASRS	[96]						
Cvpr.LScot	...L.....	[96]						
Cvpr.MMcot	...M.....	[96]						

Figure 44
Comparison of Clade C vpu Protein Sequence Reconstructions

	10	20	30	40	50	60	70	80]	
[
[
Cvpu.mrca	MLDLIARVDY	RLGVGALIVA	LIIAIVVWTI	VYIEYRKLVR	QRKIDWLIK	IRERAEDSGN	ESDGDTEELS	TLVDMGHLRL	[80]
Cvpu.LScotL....L.E.....	..M.....	[80]
Cvpu.MMcotL....L.E.....	..M.....	[80]
[
[
Cvpu.mrca	LDVNDL	[86]							
Cvpu.LScot	[86]							
Cvpu.MMcot	[86]							

Figure 45A

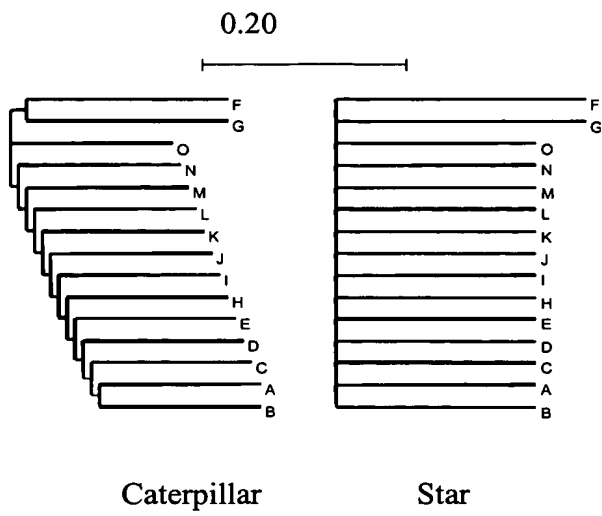


Figure 45B

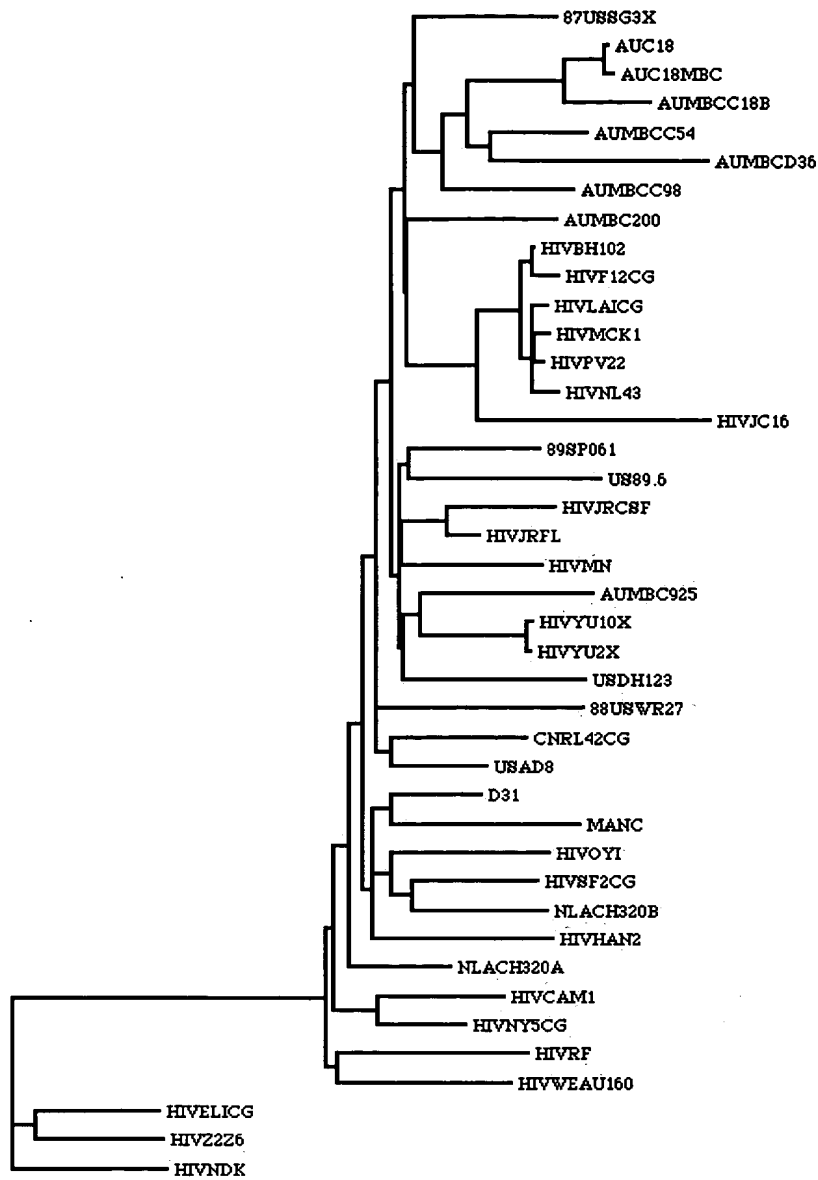


Figure 46. Deduced ancestor protein sequences

A.

SIVBK28 ancestor (Env segment)

NKSETDRWGLTKSNETSSCIAQNNCTGLEQEQMISCKFNMTGLKRDKTKEYNETW
YSTDLVCEQGNSTDNESRCYMNHCNTSVIQESCDKHYWDTIRFRYCAPPGYALLRC
NDTNYSGFMPKCSKVVSSTRMMETQTSTWFGFNGTRAENRTYIYWHGRDNRTII
SLNKYYNLTMKCRPGNKTVLPVTIMSGLVFHSQPINDRPKQAWCWFGGKWKDAI
KEVKQTIVKHPRYTGTNNTDKINLTAPGGGDPEVTFMWTNCRGEFLYCKMNWFLN
WVEDRDVTTQRPKERHRRNYVPCHIRQIINTWHKVGKNVYLPPREGDLTCNSTVTS
LIANIDWTDGNQTNITMSAEVA

B.

AN1-EnvB

MRVKGIRKNYQHLWRWGTMLLGLMLICSAAEKLWVTVYYGVPVWKEATTTLFC
ASDAKAYDTEVHNWATHACVPTDPNPQEVVLENTENFNMWKNNMVEQMHEDE
IISLWDQSLKPCVKLTPLCVTLNCTDDLRTNATNTTSSATNTTSSGGGTMEGEKG
EIKNCSFNVTTISIRDKMQKEYALFYKLDVVPIDNDNNNTNNNTSYRLINCNTSVITQ
ACPKVSFEPIPIHYCTPAGFAILKCNDKKFNGTGPCTNVSTVQCTHGIRPVVSTQLLL
NGSLAEEEVVIRSENFDTNAKTIIVQLNESVEINCTRPNNNTRKSIPIGPGRALYATGK
IIGDIRQAHCNLSRAKWNNTLKQIVTKLREQFGNNKTTIVFNQSSGGDPEIVMHSFN
CGGEFFYCNSTQLFNSTWHFNGTWGNNNTERSNNAADDNDTITLPCRIKQIINMWQ
EVGKAMYAPPISGQIRCSSNITGLLLTRDGGNNENTNNTDTEIFRPGGGDMRDNWS
ELYKYKVVKIEPLGVAPTKAKRRRVQREKSAVGMLGAMFLGFLGAAGSTMGAAS
MTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLKD
QQLLGIWGCSGKLICTTAVPWNASWSNKS�DKIWNMTWMEWEREIDNYTGLIYT
LIEESQNQQEKNEQELLELDKWASLWNWFDITNWLWYIKIFIMIVGGLVGLRIVFAV
LSIVNRVRQGYSPLSFQTHLPAPRGPDRPEGIEEEGGERDRDRSGRLVNGFLALIWD
DLRSLCLFSYHRLSDLLIVARIVELLGRRGWEALKYWWNLLQYWSQELKNSAVSL
LNATAIAVAEGTDRVIEVVQRACRAILHIPRRIRQGLERALL